

THIS WEEK



EDITORIALS

SUPERNOVAE How the biggest stars tend to go out with a bang **p.400**

WORLD VIEW Britain looks to the past for innovative inspiration **p.401**

VENOM Enzymes offer clues to the power of a spider bite **p.402**

Still much to learn about mice

A project that aims to mutate every gene in the mouse genome to improve our knowledge of mouse biology should help to avoid irreproducible results and costly failures in drug development.

The mouse is the undisputed king of laboratory science. It achieved its royal status after it was chosen as the first mammal — after the human — to have its genome sequenced. Understanding the genome made it possible to develop new molecular technologies to make mutant mice, and scientists have made them by the thousand. They have also used these mutant mice to illuminate how genes and the molecular pathways they control operate in health and disease. This has also cast some welcome, if indirect, light on human diseases.

Specialized repositories have sprung up around the world to accommodate these mutant mice and to allow them to be shared. Everybody benefits: researchers, who can have the latest mouse mutants sent to them; and science more broadly, as the repositories guarantee the quality of the genetics and the health of each strain, which is crucial for comparing the results of different experiments.

That quality must be defended. At a meeting in Munich, Germany, earlier this month, representatives of repositories from China, the United States, Europe, Japan, Canada and elsewhere expressed a concern: new technology renders it so easy to make a knockout mouse that more scientists may start to use it, without being aware of the general genetics expertise needed. It is a concern that deserves broader discussion.

At present, the ability to make a high-quality knockout of a gene in a mouse requires considerable skill in genetics and breeding techniques. But new and disruptive technologies — gene-editing methods such as CRISPR — have entered the scene, making mouse engineering considerably less challenging. But will this mean a series of strains produced with inadequate quality control? If so, experiments will be harder to reproduce, and medical research could suffer.

Since 2010, those involved with the mouse repositories, together with other geneticists, have been coordinating the International Mouse Phenotyping Consortium (IMPC). The consortium aims to make a conditional mouse mutant — in which the targeted gene can be switched off to order — for every gene in the mouse genome in a defined genetic background. Each mutant mouse will be examined in detail to find out exactly what changes occur in the animal's physiology, anatomy or behaviour when the gene is removed. It is a colossal task, with a colossal estimated price tag of US\$900 million to be shared by participating nations.

The first thousand of these phenotyped mutants will be available in a couple of months. Fifteen thousand will be available by 2021 if all goes to plan. But that plan assumes that the requisite funding will continue to flow. And like all those with power, the mouse has enemies, whose views may shake the confidence of funding agencies, already notoriously averse to large, long-term investments such as repositories. Mouse mutants are invaluable in understanding biological processes and what can go wrong in biochemical or cellular pathways in diseases such as cancer or Alzheimer's. Too often,

however, scientists consider them models of human disease, as if a manipulated gene or two could actually recapitulate a disease in a different species. Therapies that 'cure' a mutant mouse but then fail in the clinic, bring the mouse into disrepute — as recently lamented by Steven Perrin, of the Amyotrophic Lateral Sclerosis Therapy Development Institute in Cambridge, Massachusetts, who has witnessed the phenomenon too many times in relation to this disease (*Nature* **507**, 423–425; 2014).

“Therapies that ‘cure’ a mutant mouse but then fail in the clinic, bring the mouse into disrepute.”

Some scientists complain that the phenotyping approach is unreliable because when different laboratories knock out the same gene, they may see different consequences — adding to the current crisis in the reproducibility of biomedical results.

But the discrepancies are usually because mice in different labs are of different genetic strains; this makes a big difference to whether the function of a missing gene will be compensated for. Another cause of discrepancies can be viruses in the mice, which can change the way that genes are expressed. In fact, 12% of the strains submitted to one of the main mouse repositories, the Jackson Laboratory in Bar Harbor, Maine, are contaminated by pathogens.

This is why it is so desirable to have repositories that guard the health and genetic quality of the deposited mice. It is also why the IMPC is so important — by detailing the function of each gene in a standard genetic background, it will provide a necessary source of information for researchers for many decades, and help in the effort to ensure that biological results are reproducible. ■

Not on the label

A US push to flag foods as genetically engineered is hard to swallow.

The tiny US state of Vermont is no stranger to gourmands, particularly those with a fondness for its maple syrup and ice cream. On 8 May, Vermont carved out a new position in the national food scene when its governor, Peter Shumlin, signed into law a bill that requires foods on sale in the state that were made with genetically engineered ingredients to be labelled as such. It is the first such law in the country.

The law's fate is unclear: food-industry groups immediately vowed to challenge it in court. Vermont's attorney-general is readying the state's legal defence — the bill Shumlin signed included provisions to

fund these courtroom battles. It is tempting to see Vermont's move as the first success in a larger US movement that aims to limit the spread of genetically modified foods.

The Center for Food Safety, a consumer activist group in Washington DC, says that there are 35 similar food-labelling bills in the works across 16 states. That is not to say that all will follow in Vermont's footsteps. In 2012 and 2013, voters in California and Washington state defeated similar ballot measures. Vermont — the only state to boast a self-described socialist as a senator — is something of an outlier on the US political spectrum. But the interest in laws on labelling is a striking trend in a country that is the world's leading producer of genetically engineered crops.

There is plenty of precedent for such laws: more than 60 countries require genetically engineered foods to be labelled. Many of those countries grow few, if any, genetically engineered crops. The US labelling movement poses a number of logistical challenges. Navigating a patchwork nation in which labelling requirements vary from state to state is one obvious problem for the food industry. The sheer pervasiveness of genetically engineered crops in all manner of foods is another.

In 2013, such crops populated about half of US farmland. That included more than 400,000 hectares of sugar beets modified to withstand the herbicide glyphosate. By 2010, some 95% of the US sugar-beet crop was genetically engineered, and more than half of the processed sugar made in the country derives from sugar beets. Although neither the genetically engineered DNA nor protein remains in the finished product, laws proposed in some states would require that foods containing this sugar be labelled as 'genetically modified'.

And so it would go for most genetically engineered crops, which make their way onto the dinner table largely by way of processed foods. Herbicide-tolerant corn (maize) appears as the sweetener high-fructose corn syrup, and engineered soya beans are used to make the common food additive soy lecithin. Corn oil made from engineered corn is chemically no different from that made with conventionally bred corn. Yet some proposed laws would require a frozen pizza drizzled with corn oil made from genetically engineered corn to be labelled as 'genetically modified'.

The definition of that term is set to become even fuzzier as new technologies widen the array of genetic modifications available to crop breeders. Some are experimenting with 'cisgenics' — the science of modifying a crop by expressing genes plucked from related species. Methods that alter gene expression using RNA molecules are also in vogue. And advances in genetic engineering have yielded ways to precisely edit the genome, inserting genes at specific locations. These methods allow just a few letters of the DNA sequence to be changed.

"Determining the provenance of some engineered crops may be impossible."

It is a far cry from the days when genes that conferred insect resistance or herbicide tolerance were taken from a bacterium and shot near-randomly into crop genomes. Yet while regulators are deep in discussions about how to handle the new varieties of genetically engineered foods, popular conceptions of such foods seem largely unchanged.

Vermont's labelling law and many of the other proposals make no distinction: products of a crop engineered using recombinant DNA techniques to make heritable changes to the genome are to be labelled, regardless of whether that change was one that could have been produced through conventional methods such as breeding with relatives or exposing seeds to mutagens. It is also not clear whether the labelling laws could be enforced: determining the provenance of some of these engineered crops may be impossible, because the products will be indistinguishable from those made using conventional crops.

The issue of genetically engineered foods is a muddled one, and the debate surrounding them is heated. Some oppose the technology because they oppose industrialized agriculture; others worry that engineered crops could pose environmental hazards. And many consumers believe, despite evidence to the contrary, that the foods pose more health risks than those grown through conventional breeding and mutagenesis.

Researchers may understandably be hesitant to plunge into these turbulent waters. But the popular discourse around genetically engineered crops is in dire need of a scientific update. Without it, public discussion and political legislation will continue to drift away from reality. ■

Out with a bang

The discovery of a Wolf-Rayet supernova rebuts the idea that the biggest stars go quietly.

A long time ago, a faraway star threw up its insides and ended its days in a colossal explosion. The first light to hold the record of this supernova reached Earth about this time last year. Just a few hours later, quick-thinking astronomers were able to point a telescope at the hole in the sky where the star had been. The resulting images help to resolve a key question in stellar physics. And they might raise more questions about the fate of Earth.

Supernovae are one of the most stunning events in the night sky; the explosions are so well known for their violence that the term has even entered common parlance. Yet supernovae are rare, and so, therefore, are direct observations of the circumstances immediately before and after them.

As astronomers describe on page 471 of this issue, being able to focus on the immediate aftermath of a supernova has shed new light on why some stars go bang with such force. In this case, the emissions spectra sent out by the dying star show that it was a Wolf-Rayet star, massive bodies that shed their mass rapidly in strong stellar winds.

The finding is significant because, although astronomers assumed that Wolf-Rayet stars would go supernova, there was no direct evidence that they did. In fact, in the absence of observations of such

supernovae, a rival theory was gaining ground: that they might end their lives not in a bang but with a whimper. As John Eldridge explains in an accompanying News & Views article on page 431: "Until this event, there was growing evidence that such stars were likely to have dim or unobservable deaths."

Wolf-Rayet stars are more than 20 times more massive than our Sun and are very breezy places: their fierce stellar winds can reach more than 1,000 kilometres per second. They are also rare, so if the name rings a bell then it could be because you have heard of a particular specimen: WR 104, a binary star about 2,450 parsecs (8,000 light years) from Earth that shot to fame in 2008 when astronomers warned that we could be in the firing line if it exploded. If you are concerned by this (and you probably needn't be), then the finding that Wolf-Rayet stars do go supernova will do little to ease your anxiety.

A mere supernova would not threaten us at that distance, but some very massive stars explode as two powerful beams of lethal radiation known as γ -ray bursts. Depending on which way WR 104 is pointing — and the jury remains out on that — one of those bursts could head our way.

There are plenty of ifs and buts there — evidence suggests, for instance, that WR 104 has the wrong environment for γ -ray bursts — but, technically, the odds of such an event just shortened, very slightly. All Wolf-Rayet stars will go bang, the paper proposes, WR 104 included. The question is when — it could be next week, or thousands of years hence. Or it may already have happened. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



A Longitude Prize for the twenty-first century

The UK Government's new prize for substantial innovation to address pressing societal problems should be welcomed, says Martin Rees.

It is 300 years since the British government pioneered a new approach to problem-solving. The Longitude Act of 1714 offered a prize of £20,000 (worth more than £1 million today) to anyone who could devise an accurate method to determine a ship's position at sea. Among those on the committee that judged the merit of the entries was the serving Astronomer Royal.

History is repeating itself. In 2014, there is a pressing need for the United Kingdom to channel more brainpower into innovation, to jump-start new technologies and to enthuse young people. There are broad societal problems that demand fresh thinking. So Britain is reconvening the historic Longitude Committee, this time with a promised reward of £10 million (US\$17 million). Some things are different: in 2014, the challenge to be addressed by the Longitude Prize will be decided not by government officials, but by the public. And some things are the same: as Astronomer Royal, I chair the resurrected committee.

Starting on 22 May, after detailed examination by the BBC science programme *Horizon*, the public will be able to vote on the theme of the prize from a shortlist of six broad topics, each important to human welfare and each offering scope for creativity: dementia; access to safe, clean water; treating paralysis; antibiotic resistance; food and nutrition; and flight. The outcome of the vote will be announced on 25 June. An expert group will then define and set the rules for a specific challenge.

On flight, for example, the challenge for scientists and engineers could be to reduce the environmental impact of air travel. On dementia, it could be to find a way to help people with dementia to live independently for longer. The £10-million prize fund will be open for five years as a reward for the best answer to this single specific problem — although the prize could be configured (as was the challenge of 1714) so that rewards are offered for intermediate steps.

The new Longitude Prize was first announced by Prime Minister David Cameron last year. Some are likely to be sceptical of its effectiveness. I think they are wrong.

A well-designed prize should unleash investment from many quarters, amounting to much more than the prize itself, by enhancing the competitive focus on a challenge important for human welfare. The contest should also be newsworthy enough to raise the profile and reputation of innovators, and to stimulate young people's interest and enthusiasm — and that could in itself have substantial social value. For an individual or small company, the prize money is a significant incentive; for a big company, the publicity will be more important. Thus both have a motive for participating.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/7cakri

The original longitude challenge was the talk of London's coffee-houses. In the eighteenth century, Britain was a maritime power striving against Spain and the Netherlands to rule the waves. Many ships were wrecked because there was no way to keep precise track of longitude.

Astronomers believed that the answer would come from the stars. John Harrison, a working-class joiner and clockmaker with little formal education, came closest to receiving the reward money through his extraordinary mechanical talent and determination, culminating in his H4, the marine chronometer still on display at the Royal Observatory in London.

In the subsequent 300 years, the 'challenge prize' concept has been widely emulated — for instance, a prize was the stimulus for Charles Lindbergh's first transatlantic flight. In the United States, the publicly funded Defense Advanced Research Projects Agency (DARPA) has sponsored competitions for driverless vehicles. The total investment by contestants — private entrepreneurs and universities — has been much more than the \$6.5 million offered in prizes.

Pre-eminent in such ventures in the United States is the X-prize Foundation in Culver City, California, which oversees and monitors privately sponsored prizes. Its ambition is to revitalize markets that are currently "stuck due to existing failures or a commonly held belief that a solution is not possible". The first X-prize, for privately developed sub-orbital space flight, was awarded in 2004.

This type of prize has advantages over more conventional awards. The winner is decided objectively — as in athletics, and unlike the

Oscars and literary prizes. And such prizes recognize and boost up-and-coming talent — unlike the Nobel and similar prizes, for which recognition may be delayed for decades.

The original Board of Longitude lasted more than a century, and offered rewards for further discoveries and innovations. It was, in a way, the precursor of Britain's current Research Councils, Technology Strategy Boards, and so on. No longer is there a manifest number-one problem as there was in the eighteenth century. Today's research agenda is hugely more diverse and on a much larger scale; both public and private sectors provide many incentives and pathways to innovation that did not then exist. The £10-million offered by the new Longitude Prize is less than a thousandth of what Britain spends each year on research and development. But I am confident that it could have a disproportionate impact: it is surely an experiment worth trying. ■

Martin Rees is chair of the Longitude 2014 Committee.
e-mail: longitude.prize@nesta.org.uk

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Star partners form strongest magnet

Having a stellar partner might allow a dying star to become a magnetar — the strongest known magnet in the Universe — instead of a black hole.

Large stars usually form black holes when they die, so astronomers have wondered why some become magnetars, an unusual kind of neutron star. Simon Clarke of the Open University in Milton Keynes, UK, and his colleagues used the European Southern Observatory's Very Large Telescope to study the Westerlund 1 star cluster, which includes a magnetar.

They found a star that they say was probably the companion to the one that formed the magnetar. The two orbited each other closely, and as the larger one began to die, it transferred its outer layers to the smaller star. This made the small star rotate more rapidly, eventually creating an ultra-strong magnetic field.

Astron. Astrophys. 565, A90 (2014)

PALAEONTOLOGY

Oldest sperm found in fossil

Tiny fossil crustaceans found in Australia contain remarkably preserved giant sperm that are between 16 million and 23 million years old — the oldest reproductive cells ever discovered.

Renate Matzke-Karasch at Ludwig Maximilian University in Munich, Germany, and her colleagues examined ancient freshwater sediments in Queensland and used synchrotron X-rays to image the internal structure of fossil ostracods, crustaceans around 1 millimetre in length. The authors found well-preserved

detail of soft tissue, including internal organs and sperm clusters around 1.2 mm long. The team even identified nuclei in some sperm.

Many modern ostracods have huge sperm relative to their body size, and these fossils show that the trait evolved long ago, although it is not clear why.

Proc. R. Soc. B <http://doi.org/ssh> (2014)

ORGANIC CHEMISTRY

Simple recipe for small molecules

A synthesis method that uses just one chemical reaction and 12 building blocks could allow chemists to

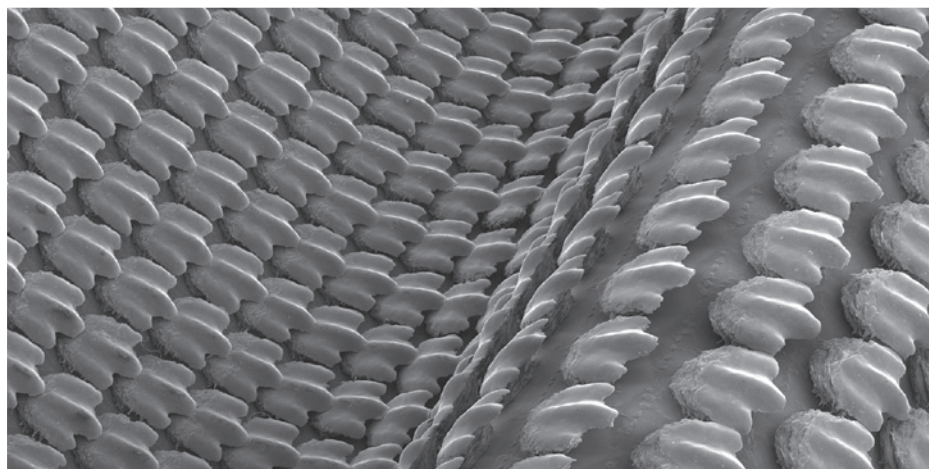
automate construction of the backbones of thousands of small molecules.

This kind of modular simplicity is standard in the laboratory synthesis of proteins, nucleotides and, increasingly, carbohydrates. Martin Burke and his colleagues at the University of Illinois at Urbana-Champaign analysed more than 2,800 natural products, including pharmaceuticals, that contain polyene motifs — chains of carbon atoms connected by alternating single and double bonds. The researchers report that more than 75% of polyene structures can be made by sequentially linking building blocks from a small library of organic acid molecules

mako shark (*Isurus oxyrinchus*). The authors compared the synthetic shark skin to a smooth control model in a robot swim test and found that the experimental skin moved 6.6% faster.

The skin eases swimming both by reducing drag and by generating vortices that boost thrust, the authors suggest.

J. Exp. Biol. 217, 1656–1666 (2014)



JAMES WEAVER/LI WEN/GEORGE LAUDER

BIOPHYSICS

Fast swimming with fake shark skin

A material that mimics shark skin enables a swimming robot to move quickly through the water by improving hydrodynamics.

Li Wen, James Weaver and George Lauder at Harvard University in Cambridge, Massachusetts, used three-dimensional printing to embed thousands of rigid tooth-like scales on a flexible membrane (pictured), based on the skin structure of the shortfin

that contain boron. These 'MIDA-boronates' were invented by Burke's group and are commercially available.

This approach avoids the need to invent a customized method for every polyene-containing compound, the authors say.

Nature Chem. <http://doi.org/ssv> (2014)

GENOMICS

Spider genomes hold venom secrets

Genome sequences from two spider species reveal the composition of their silk, and how spider venom exerts its toxic toll.

Mikkel Schierup at Aarhus



University in Denmark and his team sequenced the genomes of an African social velvet spider (*Stegodyphus mimosarum*) and a Brazilian white-knee tarantula (*Acanthoscurria geniculata*; **pictured**), and identified proteins that comprise the arachnids' venom and silk. The venom of both species contains numerous protein-cleaving enzymes that probably activate precursors of toxic venom proteins. More than 70% of velvet-spider venom is made up of enzymes that break down fats.

The team also found that the velvet spider makes a more diverse array of silk proteins than the tarantula. That is probably because it uses its silk in a greater variety of ways, such as creating complex webs to capture prey, says the team. *Nature Commun.* 5, 3765 (2014)

REGENERATIVE BIOLOGY

Stem cells make bone in monkeys

Allowing a type of stem cell to mature before transplanting it into monkeys seems to prevent the cells from forming tumours or triggering inflammation — two safety concerns that have plagued the field. Instead, the cells generated new tissue.

Induced pluripotent stem (iPS) cells can form a variety of tissues and can be made using a patient's own cells. Cynthia Dunbar at the National Institutes of Health in Bethesda, Maryland, and her team made iPS cells from the skin and bone marrow cells of two monkeys, and transplanted the cells back into the donor monkeys. These immature cells generated tumours and stimulated an inflammatory response. However, iPS cells that had developed into bone precursor cells before being transplanted did not cause

tumours or inflammation during the nearly year-long study, and formed new bone.

The findings show the promise of iPS cells for repairing tissue in humans, the researchers say.

Cell Rep. <http://doi.org/stw> (2014)

GLACIOLOGY

Antarctic area is doomed to melt

Several of Antarctica's glaciers have already begun an unstoppable meltdown, two studies suggest.

Eric Rignot at the University of California, Irvine, and his colleagues used satellite radar to measure the retreat of five glaciers in West Antarctica and found that there is nothing holding the ice sheets back from catastrophic collapse, leaving them more vulnerable than previously thought. These glaciers hold enough water to raise the global sea level by 1.2 metres.

A team led by Ian Joughin at the University of Washington in Seattle modelled the behaviour of one of these glaciers, the Thwaites Glacier, and found that it is permanently destabilized. The melting of Thwaites will probably raise sea levels by 2.5 centimetres over the next century and by more than a millimetre per year within two to nine centuries, the team says. *Geophys. Res. Lett.* <http://doi.org/srf> (2014); *Science* 344, 735–738 (2014)

MATERIALS

Plastics recycled with acid

The plastics often found in electronics typically cannot be recycled, but researchers have come up with a way of producing polymers that can be broken down into reusable monomers using acid.

Jeannette Garcia and James Hedrick at the IBM Almaden Research Center in San Jose, California, and their colleagues created a strong and stable material

SOCIAL SELECTION

Family history wins gene debate

Discussion on social media about a recent genetics talk propelled a five-year-old paper into the limelight. Yale University bioinformatician Mark Gerstein, while at the Biology of Genomes meeting in Cold Spring Harbor in early May, live-tweeted a provocative quote from Stanford University geneticist Mike Snyder's talk: "Your genome is [a] better record of your family history than your family." A lively Twitter debate ensued. Geneticist Leonid Kruglyak of the University of California, Los Angeles, following the conference hashtag from afar, tweeted a dissenting view: "Family history wins: Predicting human height by Victorian and genomic methods." He linked to a 2009 paper showing that a simple average of the parents' heights could predict a person's height better than a genetic profile of 54 locations along the genome. *Aulchenko, Y. S. et al. Eur. J. Hum. Genet.* 17, 1070–1075 (2009)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/tka69i

called a thermoset polymer, used in automotive, aerospace and electronic parts. They linked together amine-based monomers with paraformaldehyde to form the thermoset. By exposing the material to acid solutions of less than pH 2, the researchers ended up with the monomer building blocks that they could then reuse.

Science 344, 732–735 (2014)

ANTHROPOLOGY

Siberian origin for Native Americans

The discovery of a roughly 12,000-year-old human skeleton lends weight to the theory that Native Americans descended from Siberian migrants, rather than from people in a later migration.

Ancient American skulls look different from those of modern Native Americans, leading some to think that Native American ancestors arrived after humans first crossed the land bridge between Siberia and North America some 26,000 to 18,000 years ago.

James Chatters of Applied Paleoscience in Bothell, Washington, and his co-workers studied a near-intact skeleton of a teenage girl found in an underwater cave in Mexico (**pictured**). Analysis of the bones revealed features similar to those of other ancient Americans. However, DNA extracted from a tooth carried a genetic signature that occurs only in Native Americans.

The results suggest that evolution on American soil led to the differences between ancient and Native Americans. *Science* 344, 750–754 (2014)



➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Statin claims

Statements reporting that cholesterol-lowering statin drugs have adverse side effects are being withdrawn from two articles published in the *British Medical Journal*. The move was announced by Fiona Godlee, the journal's editor-in-chief, in its 17 May issue. The conclusion that around one-fifth (18–20%) of patients on statins experienced side effects was based on a “misreading” of data from a separate observational study, Godlee writes. The journal is considering whether to retract the articles fully. See go.nature.com/abswnp for more.

Brain ethics

Neuroscientists must give greater consideration to ethical issues when designing experiments, says a 12 May report from a panel tasked with advising US President Barack Obama on bioethical issues. The report is the first of two to come from the Presidential Commission for the Study of Bioethical Issues in preparation for the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative — a White House-backed project that aims to map every neuron in the human brain. The report says that scientists should consider issues such as uncovering signs of disease when scanning individual people's brains and the implications of performing neurosurgery for mental disorders.

Threats to US bees

Last winter, 23.2% of honeybee colonies in the United States perished, the US Department of Agriculture announced on 15 May. The 2013–14 death rate is lower than in previous years but is still above the level that farmers consider



FRANCISCO ZAMORA/SONORAN INSTITUTE; LIGHTHAWK

Colorado River reaches the sea

On 15 May, the Colorado River flowed all the way to the Gulf of California for the first time in 20 years. Around 130 million cubic metres of water were released from a dam on the US–Mexican border on 23 March, as part of a bold ecological experiment to quench the parched bed of the once-mighty river (see *Nature* **507**,

286–287; 2014). The water rushed through ecological restoration sites, where researchers had uprooted invasive vegetation, allowing native species such as cottonwoods and willows to germinate. The experiment is part of a 2012 water-sharing agreement between the United States and Mexico.

sustainable. Threats such as disease, insecticide exposure and changes in farming practices are thought to be contributing to declining bee populations in many parts of the world. See go.nature.com/txbt5 for more.

POLICY

Satellite regulation

The US government has relaxed restrictions on the export of satellites and their component parts to other countries. The long-awaited move, announced on 13 May, shifts most commercial, scientific and civilian satellites from a munitions list tightly

regulated by the Department of State into the hands of the Department of Commerce. University space scientists will now find it easier to collaborate with international partners. But nearly all commercial crewed spacecraft remain under the control of the old munitions regulations.

University ranking

A European consortium launched a new way to rate universities on 13 May. Known as U-Multirank, the system is a searchable database of the performance of universities from around the world. Institutions are assessed on a wide range of indicators,

such as regional research collaboration and support for students studying abroad. The system allows people to look at the performance profile of an institution; compare one university with another; and search for universities that perform well in particular areas. It differs from existing rankings, which tend to rate universities in a single league table, and rely mainly on research performance.

China open access

The Chinese Academy of Sciences and the National Natural Science Foundation of China both announced new open-access policies on

16 May. All researchers and graduate students who are supported by the agencies are now required to make their publicly funded studies free to read. Research papers must be made accessible in online repositories no later than a year after their publication in an academic journal. See go.nature.com/ojrmfy for more.

EVENTS

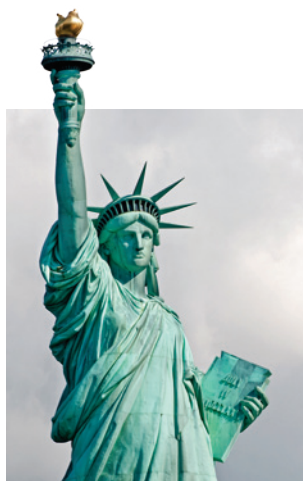
Swift slime

The single-celled winners of an unusual competition to crown the world's fastest slime mould were announced on 19 May. The Dicty World Race, held in Boston, Massachusetts, pitted strains of the slime mould *Dictyostelium discoideum* in a microfabricated racetrack 800 micrometres long. The champion was submitted by a team from the University of Groningen in the Netherlands, and was engineered to be particularly sensitive to a signalling molecule called cyclic AMP, which was used as a chemical attractant in the race. See go.nature.com/6ezbrd for more.

CULTURE

US sites at risk

Historic US landmarks are at risk of damage and destruction caused by



sea-level rise, flooding and wildfires, warns the Union of Concerned Scientists in a report released on 20 May. The advocacy group, based in Cambridge, Massachusetts, says that protecting these sites should be part of a national climate-change strategy. Among the top 30 at-risk sites are Jamestown in Virginia, the Statue of Liberty (**pictured**) and Ellis Island in New York, and coastal NASA facilities.

FACILITIES

Space ties cut

Russia is severing its space-science partnership with the United States, according to a statement given by Russian deputy prime minister Dmitry Rogozin on 13 May. The move is in retaliation against US-imposed sanctions linked to Russia's actions in Ukraine. Rogozin said that Russia will end its

long-running collaboration with the United States on the International Space Station in 2020. NASA, which depends on Russian carriers to ferry astronauts to and from the space station, says that it has not been officially notified about the move. See go.nature.com/6shdlx for more.

HAARP to close

The US government's High Frequency Active Auroral Research Program (HAARP) — a leading ionospheric research facility near Gakona, Alaska — is facing closure. The US\$250-million facility is used mainly by the Defense Advanced Research Projects Agency (DARPA) but on 14 May, the agency said that it would be ending its experiments there. The US Air Force, which manages HAARP, says that the programme will shut once DARPA concludes its work. See go.nature.com/wv69qs for more.

FUNDING

Australian budget

The Australian government has allocated Aus\$20 billion (US\$18.7 billion) to establish a Medical Research Future Fund to tackle diseases including diabetes and cancer, it announced on 13 May. But its budget also included cuts to the nation's science

COMING UP

25–30 MAY

The World Organization for Animal Health holds its general assembly in Paris. On the agenda are standards for world trade in animals and their products, and animal diseases. See go.nature.com/dgum4u

25–30 MAY

The quadrennial assembly of the Global Environment Facility will discuss expanding financing to support food security and sustainable cities. See go.nature.com/bodsub

26–28 MAY

Worldwide heads of research councils will discuss open access and how to support young researchers at the annual meeting of the Global Research Council in Beijing. See go.nature.com/ofcfna

agency — the Commonwealth Scientific and Industrial Research Organisation — and the abolition of its renewable-energy agency. See go.nature.com/jhkcvk for more.

BUSINESS

Deal spurned

The London-based drug company AstraZeneca has rejected another buyout offer from its New York rival Pfizer. On 18 May, Pfizer said that it had offered AstraZeneca around £55 (US\$93) per share — valuing the British company at roughly £69 billion. Pfizer said that this was its final offer and it would not attempt a hostile takeover. But it urged AstraZeneca's shareholders to encourage their company's board to sell.

➔ **NATURE.COM**

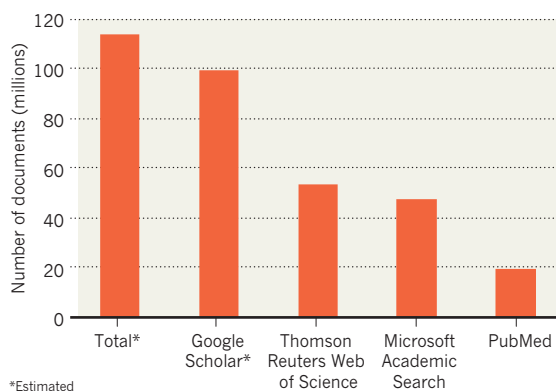
For daily news updates see: www.nature.com/news

TREND WATCH

The academic search engine Google Scholar can find about 88% of all English-language scholarly documents on the World Wide Web, according to an estimate by computer scientists Lee Giles and Madian Khabza at Pennsylvania State University in University Park (M. Khabza and C. L. Giles *PLoS ONE* 9, e93949; 2014). The duo studied the coverage of Google Scholar and a competitor, Microsoft Academic Search. At least 24% of documents are freely available, they add. See go.nature.com/matsio for more.

THE WEB OF SCHOLARSHIP

Around 114 million English-language scholarly documents (including papers, books and technical reports) can be found on the web.



NEWS IN FOCUS

OCEANOGRAPHY Researchers salvage expeditions after loss of deep-sea sub **p.408**

RESEARCH MANAGEMENT Spy agency mines text to identify next big technology **p.410**

GENOMICS Sea creature may have evolved unique nervous system **p.411**



REPRODUCTION Egg treatment mixes genes from two mothers **p.414**

JHU/APL/SWRI



The New Horizons probe (artist's impression) is set to pass Pluto and its moon Charon in July next year.

PLANETARY SCIENCE

Pluto-bound probe faces crisis

NASA scientists scramble to find an object in the outer Solar System's Kuiper belt in time for a close-up visit.

BY ALEXANDRA WITZE

Nearly 4.3 billion kilometres from Earth, and most of the way to Pluto, NASA's New Horizons spacecraft is in danger of missing out on half of its mission. Project managers face a looming deadline to identify an icy object in the outer Solar System for the probe to fly by after it passes Pluto.

A visit to a Kuiper belt object, or KBO, was always meant to be a key part of New

Horizons' US\$700-million journey, which began in 2006. But there is only a slim chance that astronomers will find a suitable KBO with their current strategy of using ground-based telescopes — and securing time on the orbiting Hubble Space Telescope is far from guaranteed.

New Horizons will fly past Pluto in July 2015. Soon afterwards, it must fire its engines and set itself on course to fly past a selected KBO. Project scientists must identify a

KBO in the next several months if they are to determine the necessary trajectory well enough for New Horizons to aim accurately and meet its target.

"They're running out of time," says Mark Sykes, director of the Planetary Science Institute in Tucson, Arizona, who is not involved in the mission. "We're not just talking about science being lost — we're talking about getting return on our investment."

New Horizons scientists have asked for 160 orbits' worth of observing time on the hugely oversubscribed Hubble. It is a rare request for a NASA mission already in operation. The committee that allocates Hubble time will make a decision by 13 June.

The two targets of New Horizons — Pluto and a KBO — each promise different rewards. "Two different scientific communities are getting excited about it," says Will Grundy, mission co-investigator at the Lowell Observatory in Flagstaff, Arizona. The Pluto fly-by will be the first close-up glimpse of this geologically active world and its moons. The KBO visit, 2–3 years later (see 'Far horizons'), would be the best look yet at one of these primordial icy bodies — time capsules from the early days of the Solar System.

In theory, project scientists should have identified a suitable KBO long ago. But they postponed their main search until 2011, waiting for all the possible KBO targets to begin converging on a narrow cone of space that New Horizons should be able to reach after its Pluto encounter. Starting to look for them before 2011 would have been impossible, says Grundy, because they would have been spread over too much of the sky.

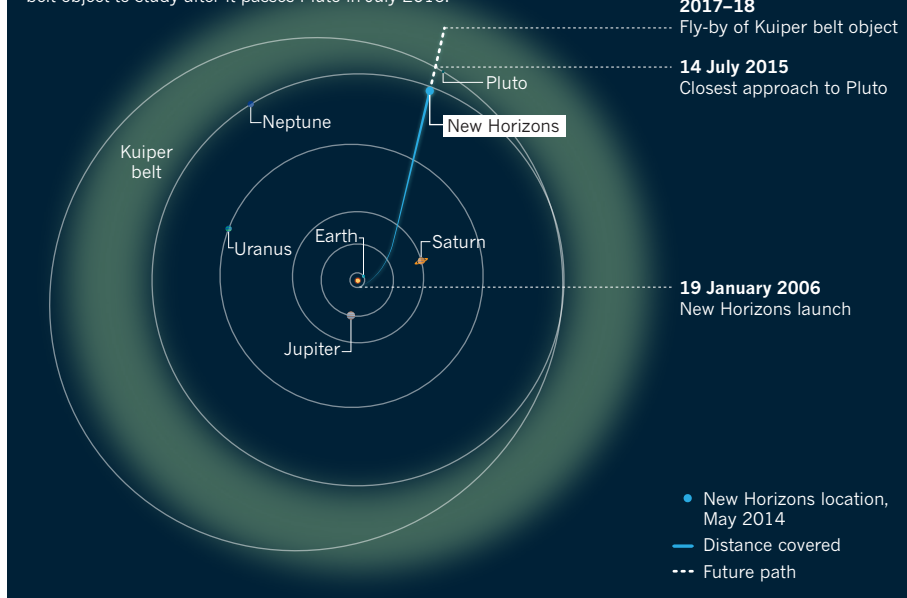
Now that the hunt for KBOs is on, the New Horizons researchers have mainly been using the 8.2-metre Subaru Telescope in Hawaii and the 6.5-metre Magellan Telescopes in Chile. They have found about 50 new KBOs; none is close enough for New Horizons to reach.

Part of the problem is that the search area is in the Galactic Plane, where most of the Milky Way's stars are. The bright stars tend to outshine the light coming from the faint KBOs, and the crowded field of view means that the discovery rate depends strongly on observing conditions. Bad weather and poor-quality observation on many nights has ruined much of the search.

Early on, the team also overestimated the number of KBOs that it would find. In ►

FAR HORIZONS

NASA's New Horizons probe has struggled to identify a Kuiper belt object to study after it passes Pluto in July 2015.



► the past couple of years, astronomers have discovered that there are fewer dim KBOs than was extrapolated on the basis of the number of bright ones. “There just aren’t as many faint ones,” says Grundy — and faint, faraway KBOs

are what New Horizons must detect.

For a mission that has been carefully planned for many years, the failure to find a KBO target is striking. But Sykes is not alarmed. “That should have been in the

calculus, but it really wouldn’t have changed their ground-based campaign,” he says.

The team has another eight nights with Subaru in late June and July, and hopes to win some more time in August. It has also applied for extra observing time with Magellan in October. Even if conditions are near perfect, the chances of snaring a KBO from the ground this year are less than 40%. With the requested Hubble time, those chances soar to greater than 90%.

Without Hubble, it might still be possible to identify a KBO target from the ground early next year. But that would complicate the mission, because New Horizons would have to wait for longer after visiting Pluto to fire its engines. In that case, the spacecraft would need to execute a sharper turn than team scientists would like to reach the correct trajectory.

In the worst-case scenario, New Horizons would be forced to observe a KBO from far away. Several possible long-distance targets have already been spotted (S. D. Benecchi *et al.* *Icarus* <http://doi.org/10.1016/j.icarus.2014.05.001>; 2014). Even with its puny 21-centimetre telescope, New Horizons would get a better glimpse of a distant KBO than the 2.4-metre Hubble can manage from an Earth orbit.

Not that the team has yet given up hope of a close KBO target. “If we can find one,” says Grundy, “we will happily take it.” ■

SOURCE: NASA/JHU/APL

OCEANOGRAPHY

Submersible loss hits research

But scientists remain positive about the future of deep-sea exploration despite disintegration of unique US Nereus craft.

BY DANIEL CRESSEY

Researchers have lost access to a huge and unexplored area of the ocean after one of the world’s most advanced submersibles blew apart 10 kilometres underwater during a deep-sea expedition. The loss of Nereus, operated by the US Woods Hole Oceanographic Institution (WHOI), leaves ongoing and future projects high and dry. But scientists who study the bottom of the ocean have told *Nature* that their research programmes will continue, and that the design of a potential successor to Nereus is already under way.

Nereus was in the Kermadec Trench in the Pacific Ocean on 9 May when its five camera feeds suddenly went blank. Researchers aboard the RV *Thomas G. Thompson* were unable to contact the US\$8-million craft. They learned its fate when pieces of its hull floated to the surface.

“It was a horrible feeling — of denial and disbelief,” says Tim Shank, a biologist at the

WHOI in Massachusetts and chief scientist on the expedition.

It is likely that one of the ceramic vessels used to encase components on Nereus collapsed under the water pressure, leading to an implosive chain reaction.

Shank’s expedition was part of a US National Science Foundation (NSF) programme to investigate the hadal zone. This area of the ocean, in deep-sea trenches between 6,000 and 11,000 metres down, is one of the least explored regions on Earth.

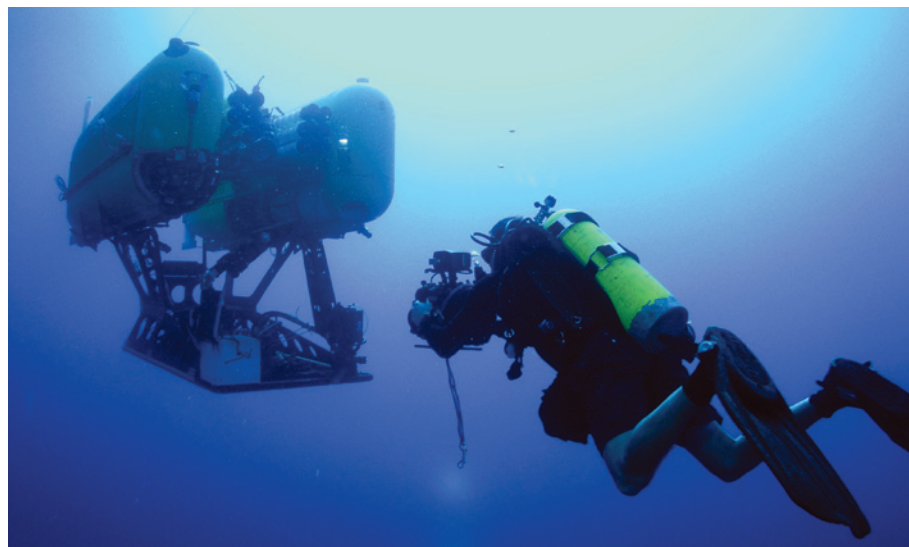
Researchers have long dismissed the idea that the area is a ‘dead zone’. For example, on its first deep dive in May 2009, Nereus discovered a new species of anemone in the deepest part of the ocean, Challenger Deep in the Pacific’s Mariana Trench. But systematic exploration has been lacking, leading in 2011 to the creation of the NSF project, an international collaboration called the Hadal Ecosystem Studies (HADES) programme. It aims to determine

the composition and distribution of hadal species, and the role of hadal pressures, food supply and trench topography on community structure.

During the final dives of Shank’s expedition, the first for HADES, researchers were measuring the metabolic rates of invertebrates such as sea cucumbers. This was done by using Nereus to place the creatures in a respirometer mounted on a separate platform. The work, which began in April, would have provided crucial data on whether trenches act as carbon sinks, and would have enabled researchers to start unpicking the food webs of the deep ocean.

“The loss of Nereus is an absolute tragedy for hadal research. We no longer have the capability to explore the lower two-thirds of our ocean with a remotely operated vehicle,” Shank told *Nature*.

Access to the deep is an ongoing challenge because only a few specialized vessels can reach this part of the ocean. The WHOI’s *Alvin*



The Nereus submersible is thought to have been lost after extreme water pressure led to an implosion.

vehicle, which is crewed rather than operated remotely, can dive to only about 4,000 metres. China, Japan, France and Russia have crewed submersibles that can reach depths of 6,000–7,000 metres (see 'Into the abyss'). Several remotely operated or autonomous research robots are also in use, but few can withstand the enormous pressures of the hadal zone.

The only other operational vessels to rival Nereus for depth range are the *DEEPSEA CHALLENGER* craft used by film director James Cameron for his 2012 voyage, and Japan's ABISMO vehicle, which has reached 10,250 metres. But *DEEPSEA CHALLENGER* is not a scientific craft, and ABISMO's equipment is not as advanced as Nereus's was.

Deep-sea data collection is still possible without Nereus. For example, Shank and his team have been able to continue some of their work by deploying other equipment, such as water and sediment samplers. Other affected expeditions are also making alternative arrangements. Nereus was due to take part in another HADES mission later this year, to explore the Mariana Trench, and expedition chief scientist Jeffrey Drazen of the University of Hawaii at Manoa says that the mission will still go ahead. Its aim is to collect animals, images and sediments from the hadal zone, and some of the work can be conducted using baited traps and instrument platforms, called

landers, that sink to the bottom, says Drazen. His team is also investigating deploying a towed camera system.

However, these are static technologies. Nereus gave real-time access and allowed scientists to point cameras directly at objects of interest and to collect samples on the fly, rather than throwing equipment off the side of their ship and hoping for the best.

There is a glimmer of hope for deep-sea exploration with submersibles. Even before the demise of Nereus, the WHOI was in discussions with the Schmidt Ocean Institute in Palo Alto, California, about constructing another 'full-ocean capable' vehicle. The collaboration is in a preliminary design phase. Scientists could be using the vehicle by mid-2016, says Victor Zikov, the Schmidt's director of research.

The WHOI is in contact with its insurers over Nereus. A payout could be used to build two of the next-generation vehicles envisaged by Schmidt, instead of one.

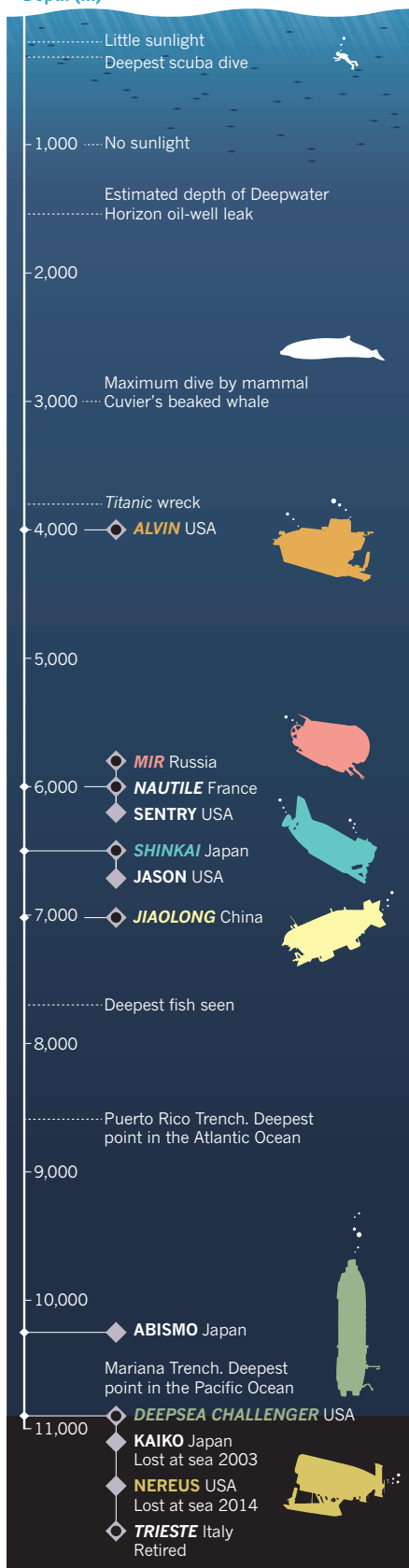
As the Nereus disaster shows, having a second submersible is a good idea, says WHOI engineer Andy Bowen, an expert in remotely operated vehicles who was lead scientist on the design of Nereus. "This is a very forceful reminder of how difficult it is to work in the ocean in general, and at the extremes even more so," he says. ■

INTO THE ABYSS

The world's fleet of deep-sea submersibles has dwindled further with the loss of Nereus.

● Crewed submersible (not to scale) ◆ Uncrewed

Depth (m)



MORE ONLINE

TOP NEWS

- Palaeoamerican skeleton is studied in underwater cave. go.nature.com/55tfch
- Gravitational-waves finding under scrutiny. go.nature.com/omv8ih
- Solar wind could trigger lightning. go.nature.com/ufjv19

NATURE PODCAST



Babies with three parents, and the sea gooseberry's nervous system. nature.com/nature/podcast



The rise of technologies such as satellite tracking could have been foretold by analysing past literature.

RICHARD NEWSTEAD/GETTY

TECHNOLOGY

Text-mining offers clues to success

US intelligence programme analyses language in patents and papers to identify next big technologies.

BY SARA REARDON

A project backed by a US intelligence agency might soon make it much easier to predict which technologies will one day become game-changers. Results revealed this week by the Intelligence Advanced Research Projects Activity (IARPA) suggest that clues in the wordings of, and relationships between, scientific papers and patents could foretell research successes.

The project, called Foresight and Understanding from Scientific Exposition (FUSE), could enable funders to pick winners, and help governments to keep an eye on 'disruptive technologies' — those that they feel might threaten national security or outpace regulations, for instance. Past examples include nanotechnologies and information technologies, such as the use of the Global Positioning System in mobile phones to allow tracking of individuals' movements. Last week, FUSE, a four-year project that started in 2011, entered its last phase: to predict the successes three to five years from now.

Although abstracts have been text-mined in the past for keywords and other clues in language, FUSE is one of the first projects to mine the whole bodies of papers and patents. So far, it has performed more than 2 million analyses of past data to pick out key advances, says Dewey Murdick, manager of the FUSE project. From these, it has identified several

hundred indicators, such as new collaborations or expressions of excitement in text, that highlight emerging areas.

"What we'd like to glean is understanding of the right combination of things that leads to success," says John Byrnes, a computer scientist at SRI International, an innovation centre in Menlo Park, California, whose team is one of three developing software for FUSE. To make predictions, his program mines text for keywords, citations and phrases that indicate authors' outlooks in scholarly papers.

One example he cites is the resolution of a technical problem that, once solved, led to what is now a mainstay of solar-panel technology. In the mid-1990s, millions of dollars were invested in research into solar panels that used aqueous solutions to convert photons into energy. Although promising at first, by 2008, the technology had been overtaken by the much more stable and effective solid-state solar panels. FUSE might have predicted the aqueous panels' demise, according to results presented by Byrnes and his team this week at the US Department of Energy's SunShot Grand Challenge Summit in Anaheim, California.

Scientific literature analysis, or 'scientometrics', is decades old. Organizations such as Thomson Reuters, an information firm

headquartered in New York, have long used these analyses to identify the most influential papers or researchers in a field. FUSE takes this further by mining millions of papers and patents in both English and Chinese, two of the most commonly used languages in scientific literature, says Murdick.

The analysis and indicators can predict whether a nascent field will become prominent or whether it is simply a source of excitement that will soon die out, says Olga Babko-Malaya, a research engineer at BAE Systems in Winchester, Massachusetts, who heads another FUSE team.

Her team uses software algorithms to analyse 'sentiment' in the natural language of papers. For instance, authors might say that their work builds on or contradicts a cited paper, or use descriptive language that expresses excitement.

The researchers also found that promising topics invent their own jargon and start using more acronyms. "Abbreviations imply acceptance by the community and are indicators of more mature technologies," says Babko-Malaya.

The changes in group collaborations in a field over time can also be predictive. FUSE researcher Lance Ramshaw at Raytheon BBN Technologies in Cambridge, Massachusetts, and his team are analysing networks between different topics, keywords and authors. He says that a new topic may be emerging when prominent authors start contributing to a group of papers that share common traits, or when alliances between collaborations shift.

Alan Porter, who specializes in technology forecasting at the Georgia Institute of Technology in Atlanta, agrees that retrospective predictions, such as the solar-panel example, are useful for modelling what companies have been doing and for tracking the history of a product. The more difficult task, he says, would be to use such a network to identify the "white spaces": areas between technology clusters that are ripe for new research.

Ideally, analysis will show patterns or tipping points that are common to success stories. Such patterns might eventually allow the project to forecast when a product might launch, or whether a drug will be approved by regulators, explains Babko-Malaya.

Although software is catching up, human analysts remain the best forecasters, says Murdick. "You can ask experts anything you want," he says.

In another project, Forecasting Science & Technology, IARPA is funding an online crowdsourcing project called SciCast, run by George Mason University in Fairfax, Virginia, and the American Association for the Advancement of Science. It aims to consult 10,000 scientists to help develop methods for generating accurate forecasts. "My personal bias is that it's a combination of people and machines that will ultimately provide the most useful value," says Murdick. ■

"Abbreviations imply acceptance by the community."

GENOMICS

Jelly genome mystery

Publication of the draft genetic sequence of a comb jelly reveals a nervous system like no other.

BY EWEN CALLAWAY

Comb jellies, or ctenophores, look like tiny disco balls and propel themselves around oceans using specialized hairs, lapping up small prey with their sticky tentacles. “They are aliens who’ve come to Earth,” says Leonid Moroz, a neuroscientist at the University of Florida in St Augustine.

The genome of the Pacific sea gooseberry (*Pleurobrachia bachei*), which Moroz and his team report online today in *Nature*, adds to the mystery of ctenophores (L. L. Moroz *et al.* *Nature* <http://dx.doi.org/10.1038/nature13400>; 2014). The sequence omits whole classes of genes found in all other animals, including genes normally involved in immunity, development and neural function. For that reason, the researchers contend that ctenophores evolved a nervous system independently.

Ctenophores have long vexed taxonomists. Their resemblance to jellyfish earned them a spot on the tree of life as a sister group to cnidarians (the phylum that includes jellyfish). On the basis of their nervous systems — which can detect light, sense prey and move musculature — many researchers had them branching off from the common ancestor of other animals after the sponges and flattened multicellular blobs known as placozoans, neither of which have a nervous system. Now armed with data showing that ctenophores lack many common genes, some scientists contend that these are the closest living relatives to the first animals.

Moroz’s team argues that the *P. bachei* genome, along with gene-expression data from other ctenophores, supports this theory. For example, microRNAs, which regulate gene expression in other animals, are completely missing from the sea gooseberry genome.

The biggest surprise, Moroz says, was the absence of many standard components of a nervous system. Nearly all known nervous systems use the same ten primary neurotransmitters; the Pacific sea gooseberry seems to employ just one or two. Moroz speculates that the organism might complete its nervous system using molecules that researchers have not yet found in this species, such as specialized protein hormones.

The uniqueness of this ctenophore’s nervous system leads Moroz and his team to argue that it must have evolved independently, after the ctenophore lineage branched off from other



Pleurobrachia bachei lacks many common genes.

animals some 500 million years ago. “Everyone thinks this kind of complexity cannot be done twice,” Moroz says. “But this organism suggests that it happens.”

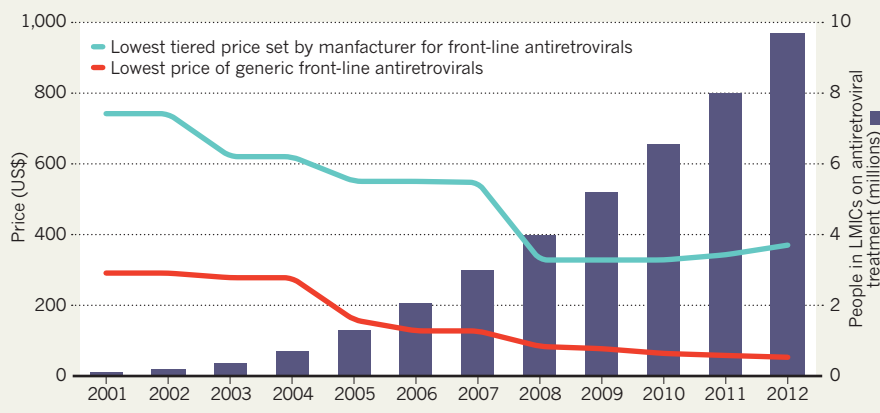
Gert Wörheide, an evolutionary geobiologist at Ludwig Maximilian University in Munich, Germany, is intrigued by the theory that the nervous system evolved twice in different animal branches, but disputes that ctenophores are the closest relatives of the first animals.

The common ancestor of all animals may have looked nothing like comb jellies, and the *P. bachei* nervous system may be a more recent adaptation, he says. “I think the last word is not spoken yet on where the ctenophores go.” ■

LEONID L. MOROZ/MATHEW CTARELLA

CUTTING COSTS

'Tiered pricing' has reduced the cost of drugs to treat HIV in low- and middle-income countries (LMICs), but bigger savings can be generated by encouraging the production of cheap generic medicines.



PUBLIC HEALTH

Activists sound alarm on tiered drug prices

Plan could increase health costs in middle-income nations.

BY ERIKA CHECK HAYDEN

Global-health campaigners and researchers are protesting against a proposal that they say would allow drug companies to raise prices for needed medicines in many developing countries.

On 13 May, a coalition of 220 non-governmental organizations (NGOs) condemned a proposal from the Global Fund to Fight AIDS, Tuberculosis and Malaria that promotes 'tiered pricing', in which drug companies offer the same medicines at vastly different prices in different countries. This approach was first used in the early 2000s to cut the cost of medications for HIV/AIDS, malaria and tuberculosis in some of the world's poorest countries. Economists say that it ultimately leads to higher costs than does allowing the manufacture of cheaper generic medicines (see 'Cutting costs').

The current debate centres on middle-income countries such as India, China, South Africa and Brazil, which have expanding economies but are home to the vast majority of the world's ill people — many of whom are poor. As these nations work to increase access to health care, they face rising prices for drugs and vaccines. But their growing overall wealth puts them at risk of losing access to prices and assistance reserved for low-income countries.

"These countries are being hit with a double whammy," says Rohit Malpani, director of policy and analysis for the Access Campaign of Médecins Sans Frontières (also known

as Doctors Without Borders) in Geneva, Switzerland.

The Global Fund, which is also based in Geneva and is the world's largest supporter of international health programmes, is trying to organize an international group to address the problem for middle-income nations; early plans for its Equitable Access Initiative (EAI) mention a global system of tiered pricing as a potential solution. According to the fund's documents, organizations that have expressed interest in the EAI include the GAVI Alliance, a vaccine initiative based in Geneva; the World Bank; the United Nations Development Programme; UNICEF, the UN children's fund; and UNITAID, a group run by the World Health Organization that negotiates medicine prices for poor countries.

Seth Faison, the Global Fund's head of communications, says that the organization aims to make drugs more affordable in needy countries. "We all have the same ultimate goal, which is maximum access to drugs at the lowest possible prices," he says.

But critics say that the fund's plan would let companies set prices that are affordable only for the wealthiest people in these areas. They point to examples such as the tiered pricing for sofosbuvir, a breakthrough hepatitis C drug marketed by Gilead Sciences. Gilead, based in Foster City, California, charges US\$900 per

treatment course in some low- and middle-income nations, such as Egypt, Mozambique, Kenya and Myanmar (see *Nature* **508**, 295–296; 2014). This is a steep discount from the \$84,000 that the company charges in the United States — where private insurers can absorb some of the cost. But the tiered price is more than the gross domestic product (GDP) per capita in Mozambique, about the same as the GDP per capita in Kenya and about one-third of that in Egypt.

"In middle-income countries with high income inequalities, companies make more money by selling to the rich people than they do by selling to the 90% of people who are poor," says Brook Baker, a senior policy analyst for the Health Global Access Project, a non-profit organization based in New York City.

Many would rather see the Global Fund contribute to or support existing access initiatives. This week's annual meeting of the World Health Organization in Geneva, for instance, includes discussions of four projects that aim to test new ways to develop affordable medicines.

The message may be getting through. The latest drafts of the Global Fund's proposal have backtracked: a version released to the fund's board on 15 May says that mentions of tiered pricing were made "in error". Yet its list of solutions to rising drug costs puts "economic classification of countries" first — leaving open the possibility that prices might be set according to a country's wealth. Middle-income countries could still be made to pay prices that are unaffordable for many of their citizens.

This alarms global-health researchers such as Suerie Moon at the Harvard School of Public Health in Boston, Massachusetts. "We have ten years of experience that shows that tiered pricing is going to lead to higher pricing," she says.

One study found that the US President's Emergency Plan for AIDS Relief, the world's largest single-country medical-aid programme, saved \$323 million between 2005 and 2008 by buying generic HIV medicines instead of tiered-priced drugs (C. B. Holmes *et al.* *J. Am. Med. Assoc.* **304**, 313–320; 2010).

Moon is also worried about the EAI's potentially far-reaching impacts. Currently, companies negotiate tiered prices with individual countries and other buyers, but a systematic global pricing scheme might force more countries to pay higher prices.

"When you have multiple agencies coming together to put in place a system of rules and policies that have not worked well in the past ten years, I think that's a real cause for concern," says Moon. ■

CORRECTION

The Editorial 'An accident waiting to happen' (*Nature* **509**, 259; 2014) stated that 13 workers were exposed to levels of contamination. In fact, 21 were exposed.



Techniques that transfer DNA from diseased human eggs to healthy ones — creating offspring with three biological parents — are on the verge of clinical use.

BY EWEN CALLAWAY



ouglass Turnbull spends much of his time seeing patients who have untreatable, often fatal, diseases. But the neurologist has rarely felt more helpless than when he met Sharon Bernardi and her young son Edward.

Bernardi had lost three children within hours of birth, owing to a mysterious build-up of acid in their blood. So it was a huge relief when Edward seemed to develop normally. “He did all his milestones: he sat up, he crawled and started to walk at 14 months,” Bernardi recalls. But when he was about two years old, he began to fall over after taking a few steps; he eventually started having seizures. In 1994, when Edward was four, he was diagnosed

ILLUSTRATION BY VASAVA

Leigh's disease, a condition that affects the central nervous system. Doctors told Sharon that her son would be lucky to reach his fifth birthday.

Turnbull, who works at Newcastle University, UK, remembers despairing that "whatever we do, we're never going to be able to help families like that". His frustration sparked a quest to develop assisted-reproduction techniques to prevent disorders such as Leigh's disease, which are caused when children inherit devastating mutations in their mitochondria, the cell's energy-making structures.

The procedures — sometimes called three-person *in vitro* fertilization (IVF) — involve transferring nuclear genetic material from the egg of a woman with mutant mitochondria into another woman's healthy egg. Turnbull and others have tested the techniques in mice, monkeys and human egg cells in culture; now, they say, it is time to try them in people. The UK Parliament is set to vote on the issue later this year; if legislation passes, the country would be the first to allow this kind of genetic modification of unborn children.

But some scientists have raised concerns over the safety of the procedures, and an increasingly vocal coalition of activists, ethicists and politicians argues that a 'yes' vote will lead down a slippery slope to designer babies. US regulators and scientists are closely watching the debate as they consider allowing similar procedures. "I admire what they've done in Britain," says Dieter Egli, a stem-cell scientist at the New York Stem Cell Foundation, a non-profit research institute. "I think they are far ahead in discussion of this, compared to the US."

A FATAL INHERITANCE

The mitochondrion, according to one popular theory, was once a free-living bacterium that became trapped in a host cell, where it boosted the cell's capacity to generate the energy-carrying molecule ATP. As a result, each mitochondrion has its own genome — but it no longer has all the genes it needs to function independently (the human mitochondrial genome, for example, has a paltry 37 genes).

Unlike the genome in the cell nucleus, which includes chromosomes from both parents, all of a person's mitochondria derive from the thousands contained in the mother's egg. For reasons still being studied, the mitochondrial genome is much less stable than the nuclear genome, accruing random DNA mutations about 1,000 times faster. As many as 1 in 5,000 children are born with diseases caused by these mutations, which affect power-hungry cells such as those in the brain and muscles. The severity of the conditions depends on the proportion of diseased mitochondria a mother passes on to her children.

Turnbull first got interested in mitochondrial disease and energy metabolism in the late 1970s, when he was working as a junior doctor on a neurology ward. A member of the Royal Air Force arrived at his clinic with a mysterious ailment: whenever he went on training runs, his muscles would suddenly give out and force him to stop. Turnbull at first suspected that the airman had a mitochondrial disease — and although he turned out to be wrong, his curiosity was piqued.

Turnbull found that the treatment options for mitochondrial diseases were limited to managing symptoms, for example by prescribing anticonvulsant drugs to ward off seizures, rather than addressing the underlying biological problem. "You see them develop a mitochondrial disease and there's bugger all you can do about it," he says. The young neurologist went on to do a PhD on the inner workings of mitochondria, and has devoted his career to understanding how they malfunction.

After Turnbull met the Bernardis in the mid-1990s, a muscle biopsy confirmed that Sharon carried mutant mitochondria. "He couldn't believe I looked so well," she says. The diagnosis helped Sharon to understand some of her health problems — and her family's. Her mother, it turned out, had lost several children, and was experiencing heart difficulties in her fifties; a cousin and other family members had also lost children. "It's been a family wiped out," says Bernardi, who lost three more babies after Edward was born. Her tragedy spurred Turnbull to seek ways to keep children from inheriting their mothers' mutant mitochondria.

Others had been thinking along similar lines. In the 1980s, embryologists working with mice had begun using 'pronuclear transfer' techniques

to investigate the developmental role of egg cells' cytoplasm. The procedures involve moving nuclear DNA from one fertilized egg to another, leaving in place most of the other contents, including the mitochondria. In 1995, researchers raised the idea that similar procedures could interrupt the transmission of mitochondrial diseases in human eggs¹.

Turnbull's laboratory began replicating the mouse research in the early 2000s, aiming to move quickly to human eggs. Working with Mary Herbert and Alison Murdoch, reproductive biologists at Newcastle University and an affiliated fertility clinic that provides IVF, they planned to start with eggs that had not been fertilized correctly and had no hope of generating a fetus.

It took 18 months to convince regulators to allow the first experiments.

"THIS ISN'T 'DESIGNER BABIES'. THIS IS ABOUT PREVENTING SERIOUS, LIFE-THREATENING, DISABLING DISEASES."

The UK Human Fertilisation and Embryology Authority (HFEA) twice denied the team's application, on the grounds that the procedures would alter the "genetic structure" of the egg — illegal under the 1990 Human Fertilisation and Embryology Act, which had established the agency. In response, the researchers argued that the term was vague and did not apply to pronuclear transfers. They filed a third appeal, this time with lawyers to argue their case, and won approval in 2005.

Around the same time, the UK Parliament began updating the 1990 law. Revised legislation came into force in 2009 and prohibited clinical application of pronuclear transfers — but it allowed for the topic to be revisited by Parliament without passing entirely new laws, pending a full airing of the scientific, regulatory and ethical issues. The law change gave the Newcastle team hope that its experiments, if successful, could one day be translated to the clinic.

A STEADY HAND

Human egg cells are one-tenth of a millimetre wide, and pronuclear transfers must be done under a microscope, in a specially designed chamber that controls temperature and air flow. It takes an expert embryologist with a steady hand: "People don't breathe when they're doing this," says Herbert.

First, a fertilized egg cell is zapped with a laser, making a hole in its membrane. Then the embryologist eases a pipette into the hole and plucks out the pronuclei, twin genetic structures that result from fertilization. Next, the researcher empties a fertilized donor egg of its genetic material and squirts the pronuclei into the hollow egg. The feat takes several minutes (see video at go.nature.com/ufatcq). If the United Kingdom approves clinical use of the procedure, the egg would then be incubated for a few days until it develops into a blastocyst of between 50 and 200 cells, which would then be transplanted into a woman's uterus.

In a paper published in May 2010, the Newcastle researchers showed² that the abnormally fertilized eggs they had been using could undergo pronuclear transfer and then develop in culture almost as well as untouched egg cells. Crucially, the transferred pronuclei brought few mitochondria with them, suggesting that a resulting embryo would largely be free of any disease-causing mutant mitochondria.

But many questions remained. Could the transfers be done efficiently enough that a woman could hope to become pregnant? Did they cause subtle molecular or genetic changes that might hinder further development or cause health problems after birth? And would the UK government ever allow them to reach the clinic?

The Newcastle team has spent the past few years looking for answers, optimizing its technique in healthy human eggs. "We're reasonably comfortable there's a chance of pregnancy with this," says Herbert. The

AN UNPLANNED EXPERIMENT

Mitochondrial-transfer pioneers

It began as a way to help a handful of patients to have babies. But fertility specialist Jacques Cohen, then at Saint Barnabas Medical Center in Livingston, New Jersey, inadvertently launched an experiment that could reveal whether mitochondrial-replacement therapies are safe to try in humans.

In the mid-1990s, Cohen was struggling to help a small number of women who were unable to conceive, although they could make enough eggs for *in vitro* fertilization. The women were not old, but their eggs were a mess — the cytoplasm around the nucleus was fragmented and littered with debris.

Cohen wondered what would happen if he added a little cytoplasm from another woman's healthy egg. He tried it in mice and it worked; so in 1997, he and his team began testing the cytoplasm-transfer technique in humans. They painstakingly 'normalized' the eggs of 33 infertile women with less than a picolitre (10^{-12} litres) of another woman's egg cytoplasm. Seventeen babies were born as a result of the procedure⁷.

Cohen knew that the transplanted cytoplasm probably contained the cellular battery packs known as mitochondria, which he had hoped would enhance embryo development. Tests reported⁸ in 2001 confirmed that at least two babies had mitochondria (which each carry 37 genes) from both their mother and the cytoplasm donor. The team was the first in the world to alter a human's genetic inheritance in this way.

The health implications for the children are unclear: studies suggest that mice with such mixed mitochondria develop hypertension and obesity in middle age⁹, and have impaired cognition — they escape from mazes more slowly than normal mice, for example¹⁰. One of the babies born as a result of the procedure was diagnosed with an autism spectrum disorder, and two further fetuses had a genetic defect known as Turner syndrome (one was miscarried, the other aborted).

The team stopped performing the procedure in 2001, when the US Food and Drug Administration said that more research was required before it could be used in humans. But nobody followed up on the 17 children, who are now teenagers.

Cohen, now lab director of Reprogenetics, a pre-implantation genetic-diagnostics company in Livingston, wants to change that. He has teamed up with researchers at Saint Barnabas for a two-phase follow-up study, including phone surveys with the families and saliva tests of the teenagers, if they are willing. The saliva will show whether the teenagers' mitochondrial genes come from both their mothers and the cytoplasm donors.

Serena Chen, a reproductive endocrinologist who joined Cohen's team in 1999 and is the project's principal investigator at Saint Barnabas, says that the timing is right for the study. Its results could make a crucial contribution to US and UK debates over related techniques aimed at helping women with mitochondrial disease to give birth to healthy babies.

"We feel like this is something that would be helpful for the other researchers in this area looking at mitochondrial disease, to provide some reassuring data that human research into this area is not unreasonable to consider," she says. **Karen Weintraub**

still-unpublished experiments have proceeded slowly, partly because healthy human eggs for experimentation are hard to come by. But Herbert says that the group has already performed more than 100 pronuclear transfers on such eggs. It also hopes to conduct safety studies to assess whether the procedures alter the transferred genome or epigenome. But such checks cannot provide complete reassurance before the leap into humans, the researchers acknowledge. "We can never say for sure that it's 100% safe," says Herbert. "It has to, at some point, go to treatment."

BATTLE LINES

In 2010, the Newcastle researchers asked the UK government to consider changing the law that prohibits them from conducting their mitochondrial-replacement procedure in humans. The request prompted a flurry of hearings, consultations and reports, involving independent scientists, bioethicists, regulators, the general public and others; another scientific review is expected in the next few weeks. But the protracted process has thrown up no major roadblocks.

"I THINK WHAT EVERYBODY WAS A LITTLE BIT UNCOMFORTABLE WITH WAS JUST HOW MUCH IS NOT KNOWN."

Nancy Lee, a senior policy adviser at the Wellcome Trust, the United Kingdom's largest biomedical-research charity, praises the review as "a good example of evidence-based policy-making and informing the public as much as is possible". The London-based charity has funded Turnbull's team to the tune of £4.4 million (US\$7.4 million), and has thrown its considerable political clout behind changing the law.

Yet some scientists argue that the procedures have not been vetted rigorously enough. Klaus Reinhardt, an evolutionary biologist at the University of Tübingen in Germany, worries about incompatibilities between the nuclear and mitochondrial genomes in individuals conceived using the procedures. Both nuclear and mitochondrial genes are needed for mitochondria to function, and it is likely that gene variants in both structures have evolved together, he says. Mitochondrial replacement in mice, fruit flies and other organisms has occasionally resulted in problems with respiration, cognition and fertility, several studies have found³. Reinhardt, who has expressed his concerns to the panel in charge of reviewing the science, questions whether there are enough safety data to go forward with clinical trials. "I don't really know how robust everything is," he says. In response, Turnbull's team casts doubt on the relevance of mitochondrial-replacement experiments that use inbred lab animals, and points out that other studies of mitochondrial replacement in mice failed to find health problems⁴.

Some critics use more emotive language. In a March Parliamentary debate and a column in *The Daily Telegraph*, Conservative Member of Parliament Jacob Rees-Mogg equated mitochondrial replacement with cloning, and said that the techniques would promote eugenics. "In a country nervous about genetically modified crops, we are making the foolhardy move to genetically modified babies," he said in the debate.

An international coalition of several dozen scholars and bioethicists, many at religious institutions, expressed similar sentiments in March 2013 in a letter to *The Times* newspaper, arguing that mitochondrial replacement "would open the door to further genetic alterations of human beings with unforeseeable consequences".

To counter this opposition, Turnbull and other supporters point out that the techniques will be used only to prevent serious mitochondrial diseases. The researchers have made patients' stories, such as the plight of the Bernardi family, central to

➔ NATURE.COM

For a video of pronuclear transfer and a podcast: go.nature.com/ufatqc

their appeals. They have compared mitochondrial replacement to changing the batteries in a camera (a poor analogy, some other scientists say), and they argue that mitochondrial DNA makes up a tiny fraction of the overall genome, with little influence over a person's defining traits. "This is not a slippery slope, in my view," Turnbull says. "This isn't 'designer babies'. This is about preventing serious, life-threatening, disabling diseases."

MONKEY TRIAL

A similar debate is shaping up across the Atlantic. While Turnbull and his team were developing their pronuclear-transfer technique in human egg cells, a US team was testing a related method in monkeys. In 2009, reproductive biologist Shoukhrat Mitalipov at the Oregon Health and Science University in Beaverton and his colleagues reported the birth of two healthy rhesus macaques whose mitochondria and nuclei had come from different egg cells⁵. The monkey twins — named Mito and Tracker, after a reagent used to make mitochondria glow — were conceived through a method called maternal spindle transfer (see 'Genome transplant'). This involves shuttling an egg's nuclear genetic material to an empty donor egg before fertilization, rather than after as in pronuclear transfer. There have not yet been any side-by-side experiments to compare the merits of the two techniques, although both teams are keen to try.

Mitalipov's team has used maternal spindle transfer to conceive five monkeys, including one from a previously frozen egg (to mimic a likely clinical situation). Mito, Tracker and two others born in 2009 have celebrated their fifth birthdays, and are still healthy. Mitalipov plans to breed them soon to determine their fertility. His team has also proved its technique in human eggs: the embryos formed blastocysts, albeit at a low rate, and produced embryonic stem cells with the potential to give rise to all the body's different tissues⁶. In unpublished work, the researchers have since drastically improved the efficiency of the procedure, he says. "Now we want to transplant these embryos."

First, he will need approval from the US Food and Drug Administration (FDA). The agency has required researchers to seek permission for mitochondrial transfers since 2001, after a New Jersey fertility clinic carried out dozens of procedures that involved moving small amounts of cytoplasm — including some mitochondria — between human eggs to improve conception rates (see 'An unplanned experiment').

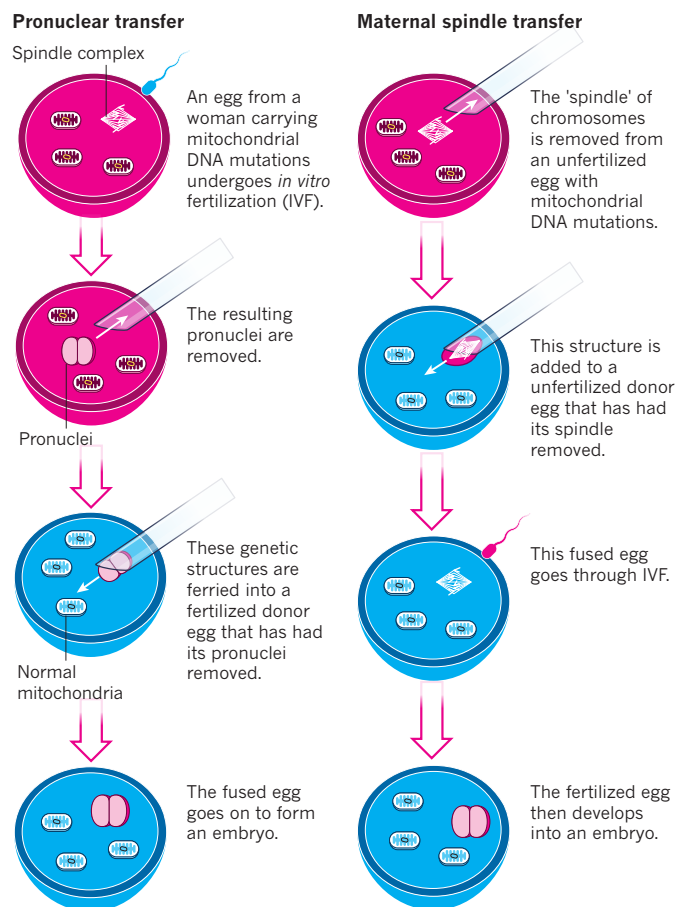
Mitalipov last year put in a proposal to carry out clinical trials in humans, and an FDA advisory panel met in February to discuss the issue. The committee spent two days chewing over the same questions that the United Kingdom has been grappling with, such as how to establish the safety and effectiveness of the procedures in cells and animal models, and what the first patient trials might look like. Committee chair Evan Snyder, a stem-cell biologist at Sanford-Burnham Medical Research Institute in La Jolla, California, says that most of his colleagues are disposed to take cellular therapies to patients, and that they recognize the potential to unshackle families from the consequences of mitochondrial mutations. But, he says, "I think what everybody was a little bit uncomfortable with was just how much is not known". Some panel members wanted to see multiple generations of monkeys born healthy using the procedures, as well as more safety work on human eggs.

Mitalipov found the meeting frustrating: "I don't want to go back and do another decade or two decades of research, which we can do. But meanwhile, there will be thousands and thousands of children born every year that will suffer." He says he would consider moving his lab to Britain to help bring his research to patients more quickly. Snyder, however, senses that his committee is not far from green-lighting clinical trials, and that safety hurdles could be surmounted in two or three years.

Back in the United Kingdom, the legislation to allow mitochondrial replacement is still being hammered out — a consultation of the draft law finishes on 21 May — but proponents are quietly confident that Parliament will say yes. The move has support across the political spectrum, and most of the scientific and ethical advice given to the government has been encouraging. However, a law change would merely give the HFEA the power to allow the procedures, and the agency would probably want more safety and effectiveness data before it approved any trials.

GENOME TRANSPLANT

Two different techniques could be used to prevent children from inheriting their mothers' mutant mitochondria.



Bernardi hopes that clinical trials will eventually go ahead. But "I think it would be bittersweet if somebody had a baby" conceived with the procedures, she adds. Her son Edward lived well beyond the expectations, although he was eventually confined to a wheelchair and his health worsened in waves as doctors struggled to find medications to quell symptoms including spasms that rendered his arms stiff and immobile. Bernardi strove to give him a normal life: he attended school, went on class trips and developed crushes. "He liked his girls, he did," she says. Bernardi resisted feeding him through a tube until he was unable to eat normally, at the age of 20. "Up until the last ten weeks, I would say he had a very good quality of life," she says.

Edward Bernardi died in March 2011 after a 21-year struggle with Leigh's disease. "I don't think this would benefit me," says Bernardi of the procedures that may be on the cusp of helping other women with mitochondrial disease. "But this keeps Edward's legacy." ■

Ewen Callaway is a senior reporter for Nature in London.

1. Rubenstein, D. S., Thomasma, D. C., Schon, E. A. & Zinaman, M. J. *Camb. Q. Healthcare Ethics* **4**, 316–339 (1995).
2. Craven, L. *et al. Nature* **465**, 82–85 (2010).
3. Reinhardt, K., Dowling, D. K. & Morrow, E. H. *Science* **341**, 1345–1346 (2013).
4. Chinnery, P. F. *et al. PLoS Genet.* **10**, e1004315 (2014).
5. Tachibana, M. *et al. Nature* **461**, 367–372 (2009).
6. Tachibana, M. *et al. Nature* **493**, 627–631 (2013).
7. Barritt, J. A., Willadsen, S., Brenner, C. & Cohen, J. *Hum. Reprod. Update* **7**, 428–435 (2001).
8. Barritt, J. A., Brenner, C., Malter, H. E. & Cohen, J. *Hum. Reprod.* **16**, 513–516 (2001).
9. Acton, B. M., Lai, I., Shang, X., Jurisicova, A. & Casoor, R. F. *Biol. Reprod.* **77**, 569–576 (2007).
10. Sharpley, M. S. *et al. Cell* **151**, 333–343 (2012).

A rash of road construction is causing widespread change in the world's largest tropical forest — with potentially global consequences.

BY BARBARA FRASER

Carving up the AMAZON



Next to a newly paved highway in the Peruvian Amazon, a discreet white-on-green sign urges travellers to protect the surrounding ecosystem. “Let’s care for the environment, let’s conserve the forest,” it reads. But the appeal comes too late for this spot in the region known as Madre de Dios. Before the route was paved a few years ago, tall trees lined the roadside, but the forest edge here now lies about half a kilometre away, beyond a jumble of underbrush and freshly cut trees where a cattle pasture was recently carved out of the woods.

As drivers head east and enter Brazil, the view is much the same for hundreds of kilometres. Such is the impact of the Inter-oceanic Highway, a route some 5,500 kilometres long that cuts clear across South America.

The highway is just one strand in a web of roads that now criss-cross the Amazon. So far, most have encroached on forest around the edges of the basin, but they are increasingly slicing through the middle. In Brazil alone, the Amazon road system grew by an average of almost 17,000 kilometres a year between 2004 and 2007 (ref. 1). Across the basin, estimates for the total length of roads vary widely from about 100,000 to 190,000 kilometres of paved and dirt roads cutting through the Amazon.

Once construction begins, road crews are quickly followed by land speculators, loggers, farmers, ranchers, gold miners and others who carve away the forest along the route. That activity leaves obvious scars on the landscape in the

form of treeless expanses, but research is now showing that the building of roads also triggers a cascade of environmental changes in the remaining forest that can dry out trees, set the stage for wildfires and weaken the ecosystem.

“Put a road into a frontier area and it opens a Pandora’s box,” says biologist William Laurance of the Centre for Tropical Environmental & Sustainability Science at James Cook University in Cairns, Australia.

The drying brought about by roads influences local atmospheric circulation patterns and can have farther-reaching effects that not only compromise the health of the Amazon but can also contribute to global warming by releasing carbon stored in the forest. Understanding those details is crucial, researchers say, for determining whether these effects — combined with severe droughts such as those that struck parts of the Amazon basin in 2005, 2007 and 2010 — could tip the world’s largest expanse of tropical forest from being a net absorber of carbon dioxide to a net emitter².

THE FIRST CUT

It was a road that kicked off the pattern of destruction in the Amazon forest. In the 1970s, Brazil began building the Trans-Amazonian Highway from near the country’s easternmost point on the Atlantic coast to its western border, where the state of Amazonas meets Peru. The route opened up the heart of the Amazon to logging, ranching and settlement, causing deforestation rates to soar. Extreme spells in the 1990s and early 2000s claimed more than 25,000

square kilometres a year — an area bigger than New Jersey. Since 2005, government measures, including crackdowns on illegal logging, have slowed forest loss. Throughout, roads have provided the means to penetrate the forest and erase large chunks of it. In an unpublished study of the Brazilian Amazon, Christopher Barber, a researcher at South Dakota State University in Brookings, found that 95% of deforestation in the region occurs within 7 kilometres of a road. And that is not the only problem: just as serious as outright deforestation is fragmentation, which happens when loggers, ranchers and farmers move in. In Brazil, up to 38,000 kilometres of new forest edge are created each year³.

Standing in a field in the western Brazilian state of Mato Grosso, Michael Coe can feel the difference that deforestation makes in the Amazon. An atmospheric scientist who heads the Amazon programme of the Woods Hole Research Center in Falmouth, Massachusetts, Coe is visiting an 80,000-hectare patch of former forest that was originally cleared some years ago to build a cattle ranch, which later morphed into a soya-bean farm. The air is noticeably hotter and drier in the field than in one of the few patches of forest left on the farm.

Coe and his colleagues are here to study how forest degradation and fires alter the flow of water and energy in the Amazon ecosystem. Evapotranspiration from trees provides moisture to the air and feeds much of the precipitation in the Amazon: when the trees disappear, so does a major source of moisture. A study using satellite data and models of atmospheric

AFP/GETTY



The Interoceanic Highway has spurred deforestation in the Peruvian Amazon.

The flames damaged more than a quarter of a million hectares of forest in that state alone and caused US\$100 million in damages. Smoke blanketed Rio Branco, the state capital, and public-health concerns finally led to ordinances to control burning during times of drought.

Scientists considered the 2005 drought to be a once-in-a-century event; some 70 million hectares of forest suffered water stress⁷, and there was significant drying within the tree canopy. But five years later, a similar dry spell struck, triggering another extreme bout of fires. Because they have not evolved in an environment frequently beset by fires, trees in the Amazon forest are susceptible to heat and damage from flames.

Farther east, in Brazil's Xingu region, researchers saw similar results from experimental fires during a drought in 2007. Tree mortality from heat and fire damage that year was more than four times that of a normal year⁸, especially along the forest edge, which the researchers burn every three years in a cycle emulating traditional Amazonian agricultural practices, says ecologist Paulo Monteiro Brando of Brazil's Amazon Environmental Research Institute in Brasília. In the Amazon, burning is the cheapest and most effective way for farmers to clear fields and give them a nutrient boost before planting crops, or rid them of ticks that plague livestock.

Understanding the future of the Amazon means learning how to model not just physical and atmospheric processes, but also how humans are changing the land, researchers say.

And as the wider impact of Amazonian roads becomes clearer, planners and conservationists face a dilemma. Although roads threaten the forest's health, they also significantly lower costs for farmers and businesses, and can make a difference between life and death for people in remote areas far from hospitals.

But unrestricted road building could lead to irreparable environmental harm, say researchers. "We're looking at a tidal wave of road expansion happening in the next few decades," Laurance says. "It's ecological Armageddon, and it's happening again and again." ■

Barbara Fraser is a freelance writer in Lima, Peru.

1. Ahmed, S. E., Souza, C. M. Jr, Riberio, J. & Ewers, R. M. *Reg. Environ. Change* **13**, 927–937 (2013).
2. Gatti, L. V. et al. *Nature* **506**, 76–80 (2014).
3. Cochrane, M. A. & Laurance, W. F. *Ambio* **37**, 522–527 (2008).
4. Spracklen, D. V., Arnold, S. R. & Taylor, C. M. *Nature* **489**, 282–285 (2012).
5. Stickler, C. M. et al. *Proc. Natl Acad. Sci. USA* **110**, 9601–9606 (2013).
6. Briant, G., Gond, V. & Laurance, S. G. W. *Biol. Conserv.* **143**, 2763–2769 (2010).
7. Saatchi, S. et al. *Proc. Natl Acad. Sci. USA* **110**, 565–570 (2013).
8. Brando, P. M. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1305499111> (2014).

“We’re looking at a tidal wave of road expansion happening in the next few decades.”

circulation suggests that air passing over tropical regions rich in vegetation produces at least twice as much rain as air moving over areas with little vegetation⁴.

Stripping away trees not only eliminates a source of moisture; it also changes the regional air flow. Heat rising from a bare field creates a low-pressure system that pulls in air from the surrounding area, sucking moisture out of the nearby forest, says Coe.

As the forest dries, it transfers less moisture to the atmosphere, changing rainfall patterns hundreds or thousands of kilometres downwind. That could affect not only forests and agriculture across the basin, but also the amount of water available to power hydroelectric dams. In a simulation using climate, hydrological and land-use models, Coe and his colleagues projected that reductions in rainfall caused by deforestation could drastically cut the power-generating capacity of Amazonian dams⁵. That would upset the plans of Brazil, Peru and Ecuador, which intend to increase hydropower to meet rapidly growing electricity demands.

The drying effect reaches well past the forest's edge. And the more fragmented the forest, the wider the impact, according to one study that found canopy drying 2.7 kilometres from the edge of a highly fragmented forest⁶.

The influence of roads in the Amazon could

even reach around the world. Recent lines of research suggest that changes in several factors prevent trees in disrupted forests from storing as much carbon as they did in the past, a shift that could accelerate global warming.

Greg Asner, a tropical ecologist at the Carnegie Institution for Science at Stanford University in California, studies the chemistry of the tree canopy in the Amazon using ground plots and airborne spectrometers. He is finding that the forest canopy along the edges of open patches does not seem to hold as much water or pigment, such as chlorophyll, as trees in unbroken parts of the forest. “Not enough chlorophyll and not enough water keep the canopy from soaking up carbon dioxide at the rate that we know it can, compared to the more interior forest,” he says.

LINE OF FIRE

Changes in the Amazon's fire potential are also impeding the forest's ability to store carbon. Conventional wisdom has long held that the rainforest was too humid to burn. But in 2005, when drought struck the western Amazon, wildfires in the Brazilian state of Acre merged into a line 11 kilometres long, with flames leaping to canopy height, recalls Foster Brown, an environmental geochemist at the Woods Hole Research Center who witnessed the fires.

COMMENT

EVOLUTION Why nineteenth century US feminism embraced Darwin **p.424**



INFORMATION How the ideas of a Belgian librarian presaged the Internet **p.425**

ENVIRONMENT Evidence-based policy needed for toxic waste in Naples **p.427**

POLLUTION Analyse Europe's spring smog to inform climate policy **p.427**

ILLUSTRATION BY RICHARD WILKINSON



A road map for suicide research and prevention

It is time for policy-makers, funders, researchers and clinicians to tackle high suicide rates, say
André Aleman and Damiaan Denys.

According to the World Health Organization, almost 1 million people kill themselves every year. That is more than the number that die in homicides and war combined. A further 10 million to 20 million people attempt it.

Suicide is one of the three leading causes of death in the economically most productive age group — those aged 15–44 years — and rates have risen since the economic crisis triggered by the banking crash in 2008 (see ‘Suicide rates in Europe’). For example, the number of suicides per year in the Netherlands rose by 30% between 2008 and 2012, from 1,353 to 1,753. In the United States, the average suicide costs society US\$1.06 million according to the US Centers for Disease Control and Prevention.

Despite its enormous societal impact, little progress has been made in the scientific understanding or treatment of suicidal behaviour. We do know that up to 90% of suicides occur in people with a clinically diagnosable psychiatric disorder¹. Large epidemiological studies have shown mental disorders, particularly depression and alcohol addiction, to be major risk factors². And there is compelling evidence that adequate prevention and treatment of such disorders can reduce suicide rates³.

But psychiatry has long neglected the topic. Other than as symptoms of borderline personality disorder and mood disorders, suicide, suicide attempts and suicidal thoughts were not listed in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. The *DSM-5* (published last year) does not code suicidal behaviour — the most prominent emergency in psychiatry in primary care. Suicidality is perceived as a medical complication rather than as a disorder in its own right.

For every study on suicidality published in the two highest-ranking general psychiatry journals (*American Journal of Psychiatry* and *JAMA Psychiatry*) over the past five years, there were six papers on schizophrenia, the incidence of which is one-quarter that of suicidal behaviour. And, in contrast to the studies on schizophrenia, those on suicidal behaviour are mostly epidemiological and do not investigate underlying mechanisms.

The lack of suicide research may be due to several factors. The first is cultural taboos. People are hesitant to talk about the suicide ►

► of a family member or friend, and many religions consider suicide dishonourable. It is often deemed unlawful, too. In India and Singapore, for example, attempted suicide is punishable by up to a year in prison. In several US states, suicide is still considered an unwritten 'common law' crime, which can have financial consequences for the family. Assisted suicide is illegal in many countries.

Second, the triggers of suicide are complicated, involving mental-health, financial, social, cultural and moral issues. Third, suicidal behaviour could be difficult to study if non-fatal suicidal attempts differ in aetiology from fatal ones.

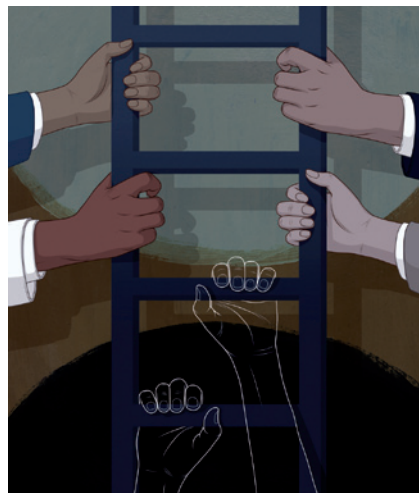
What is urgently needed is a road map for the systematic study of the mechanisms of suicidal behaviour, independent of any associated disorder. Only then can evidence-based prevention programmes be framed.

In our view, these four steps are needed.

FOUR-POINT PLAN

Define suicide as a distinct disorder. Treatment of the mental disorders commonly associated with suicide, such as depression, fails to prevent suicidal behaviour in most people. Although suicidality touches different medical and psychosocial disciplines, it should come under the remit of psychiatry, which being at the crossroads of mind and brain is well positioned to appraise all dimensions. Psychiatry should take responsibility for defining suicidality adequately, incorporating it in classification systems, developing ratings scales to predict and assess severity and examining treatment options. This will make suicidal behaviour visible as a mental disorder.

Understand the mechanisms. The roots of suicide — psychological and neurobiological



— could stem from difficulties in emotion regulation and the underlying brain circuits. The most important psychological correlates are anxiety, reduced impulse control and increased aggression². Furthermore, people showing suicidal behaviour tend to suppress their emotions and have difficulties identifying their feelings⁴. Suicidality is associated with hopelessness, sensitivity towards social disapproval and a reduced ability to imagine positive future events⁵. But there are probably several routes to suicidal behaviour.

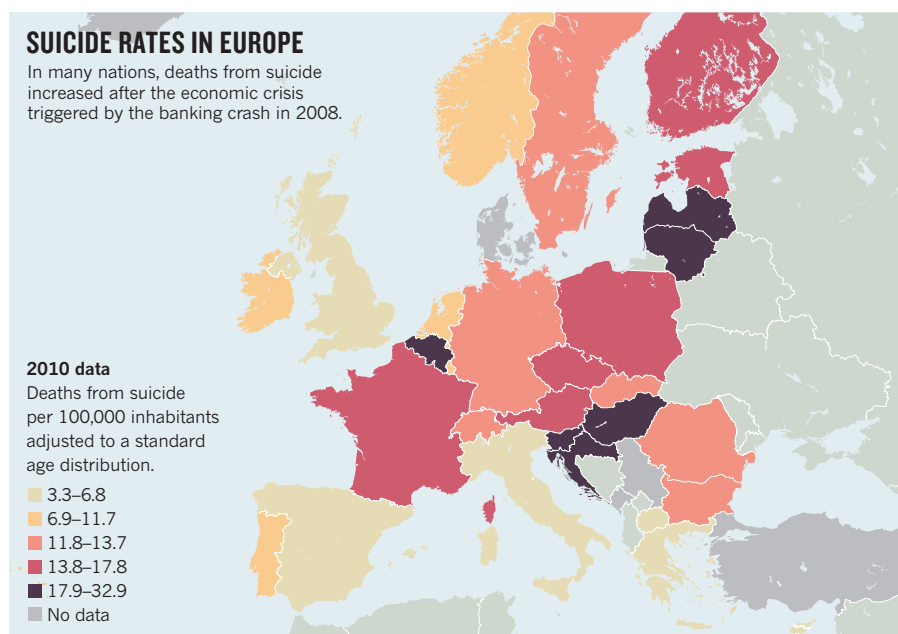
Research should focus on individual differences in cognitive control of emotion. Some people may have excessively strong emotional reactions to challenging events such as bereavement or job loss, some may lack cognitive flexibility and coping skills in the face of adversity, and some may show impulsive aggressive tendencies. Widely accepted models of the role of emotional reactivity and lack of cognitive control in setting the stage for suicidal behaviour need rigorous investigation⁶.

An example of the sort of research that we need more of is a neuroimaging study⁷ by a team led by psychiatrist Scott Matthews at the VA San Diego Healthcare System in California. They compared brain activity in combat-exposed war veterans considered to be at risk of suicide with that in those not deemed to be at risk. Both groups had similar levels of depression and post-traumatic stress disorder. Members of the suicidal group showed stronger activation of the anterior cingulate and prefrontal cortex when they made mistakes while performing a concentration task. These brain areas are involved in cognitive control and monitoring of actions. The authors hypothesize that the extra effort used to process errors during self-monitoring may represent a vulnerability that could dent people's ability to cope with stress. This deserves investigation in larger samples, taking into account different emotional states.

Fund suicide research. Governments and funding agencies should invest more in the topic. The Horizon 2020 European Union Framework Programme for Research and Innovation should include a challenge devoted to suicide research — most urgently, the definition of valid criteria for suicidal behaviour as a mental disorder and the investigation of putative abnormalities in emotion-regulation circuits associated with such behaviour. The Societal Challenges presently defined in the framework do not mention suicide. The US National Institute of Mental Health in Bethesda, Maryland, has requested applications to develop ways to screen adolescents for suicide risk, but larger and more comprehensive programmes are necessary.

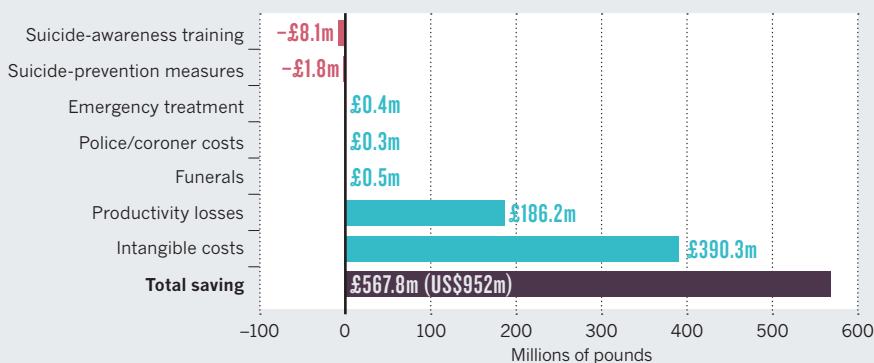
Promising in this regard is the institute's Research Domain Criteria project, which offers funding for the development of ways to classify psychopathology according to observable behaviour and neurobiological measures. For example, apathy is frequently observed in psychiatric and neurological disorders such as depression, schizophrenia, Parkinson's disease and Alzheimer's disease. Because the mechanisms leading to apathy may well be the same across these disorders, apathy is increasingly being studied as a distinct syndrome, irrespective of whether the patient has other symptoms of psychiatric or neurological disorders. Because suicide is a risk in various mental disorders, and encompasses neurobiological as well as social aspects, it is similarly suitable as topic within this project.

Promote prevention. Governments should invest as much in suicide prevention as they do in reducing fatal road accidents. In 2008–09, UK spending on road-safety awareness, including television advertisements, topped £19 million (US\$32 million); by contrast, £1.5 million was invested over



PREVENTION PAYS

Economic modelling predicts that in one year, the costs of training UK general practitioners in suicide prevention, and the ensuing costs of psychological and pharmaceutical therapy they may prescribe to at-risk people, are outweighed by the savings to the public purse owing to the roughly 600 deaths averted.



three years in suicide research. Fatal road accidents have declined steadily over the past decades, whereas suicide rates have levelled or even increased. Suicide awareness and prevention was highlighted in a review published earlier this year, which concluded that there is a return on investment for several mental-health promotion and illness-prevention interventions⁸ (see 'Prevention pays').

Risk factors are known from epidemiological studies, which should help to shape developing programmes aimed at prevention.

"Governments should invest as much in suicide prevention as they do in reducing fatal road accidents."

Notable risk factors are mental disorder, previous suicide attempt, anxiety, impulsivity in combination with aggressive tendencies, family history of suicide and stressful life events such as job loss or divorce. Comprehensive prevention programmes should be developed that incorporate state-of-the-art knowledge³.

A good prevention programme would increase awareness and mental-health literacy in the general population to improve people's understanding of warning signs. Better education is also important for general practitioners (GPs), because many people with suicidal thoughts contact their GP in the weeks before attempting suicide. Prevention programmes would offer clear and easy access points for help, and a monitoring service for those at risk. Programmes must enlist governments and other stakeholders to tackle stigma, a major obstacle to suicide prevention.

Few have systematically implemented such programmes. Examples include Finland, Scotland and the US military. Their efforts should now be evaluated to pave the way for evidence-based improvements.

Coordinated efforts are clearly needed from public-health authorities, clinicians and scientists to understand and prevent suicide. Researchers should take advantage of

progress in neurobiology and neuroimaging technology to uncover the brain mechanisms involved. Clinicians must focus on suicidal behaviour as a target of treatment in its own right. As Australia's National Mental Health Commission put it⁹: "We can and must do better." ■

André Aleman is professor of cognitive neuropsychiatry at the University of Groningen, the Netherlands. **Damiaa Denys** is professor of psychiatry at the University of Amsterdam, and is at the Netherlands Institute for Neuroscience, Amsterdam, the Netherlands.
e-mails: a.aleman@umcg.nl;
d.denys@amc.uva.nl

1. Sadock, B. J., Sadock, V. A. & Ruiz, P. (eds) *Kaplan and Sadock's Comprehensive Textbook of Psychiatry* 7th edn 2031–2040 (Lippincott Williams & Wilkins, 2001).
2. Nock, M. K. et al. *PLoS Med.* **6**, e1000123 (2009).
3. World Health Organization *Public Health Action for the Prevention of Suicide: A Framework* (WHO, 2012).
4. Pisani, A. R. et al. *J. Youth Adolesc.* **42**, 807–820 (2013).
5. van Heeringen, C., Bijttebier, S. & Godfrin, K. *Neurosci. Biobehav. Rev.* **35**, 688–698 (2011).
6. Dour, H. J., Cha, C. B. & Nock, M. K. *Behav. Res. Ther.* **49**, 294–298 (2011).
7. Matthews, S., Spadoni, A., Knox, K., Strigo, I. & Simmons, A. *Psychosom. Med.* **74**, 471–475 (2012).
8. Research Prioritization Task Force *A Prioritized Research Agenda for Suicide Prevention: An Action Plan to Save Lives* (Action Alliance for Suicide Prevention, 2014).
9. Lourey, C., Plumb, J. & Mills, A. *A Contributing Life: The 2013 National Report Card on Mental Health and Suicide Prevention* (National Mental Health Commission, 2013).

CORRECTION

The Comment 'Realizing China's urban dream' (*Nature* **509**, 158–160; 2014) wrongly stated that more than 75.8 million hectares of arable land in China could be realized by optimizing rural residential land use. The correct figure is 7.58 million hectares.



Women's advocates (left to right) Elizabeth Cady Stanton, Antoinette Brown Blackwell and Maria Mitchell.

EVOLUTIONARY BIOLOGY

Darwin and the women

Sarah S. Richardson relishes a study of how nineteenth-century US feminists used the biologist's ideas.

Two misplaced narratives dominate thinking on the historical relationship between feminism and evolutionary biology. The first is that nineteenth-century Darwinists presented a chorus of sexist views of women. The second is that feminism and evolutionary biology are wholly independent intellectual movements. In *From Eve to Evolution*, historian Kimberly Hamlin counters these misconceptions with the most comprehensive account so far of how nineteenth-century US men and women appropriated Darwinian ideas to argue for the equality of the sexes in the domestic and public spheres.

The US women's movement gathered fresh energy in the decades following the end of the civil war in 1865, launching calls for women's suffrage, access to property rights and education, and the freedom to divorce. Hamlin shows how prominent women's rights advocates enlisted science "as a force for positive change", even when excluded from lecture halls, as they often were. She demonstrates that evolutionary science offered US feminists a fresh intellectual framework from which to challenge the biblical dogma that stipulated women's inferiority to men and submissive role in domestic life. For social activists such as Elizabeth Cady Stanton, evolutionary theory "provided a new way for women to view the universe and their role in it, and a new language to describe what they saw".

As feminists rallied to respond to anti-woman screeds such as Harvard-trained



From Eve to Evolution: Darwin, Science, and Women's Rights in Gilded Age America

KIMBERLY A. HAMLIN
University of Chicago
Press: 2014.

Jacobi, astronomer Maria Mitchell and author and suffragist Helen Hamilton Gardener, argued for the importance of training women in science and for a science of sex free of misogynistic bias. Gardener's brain, left to science to prove the equality of male and female intellect, still stands on display at Cornell University in Ithaca, New York.

Other US feminists summoned the evidence of biology to argue for more-egalitarian marriage and child-rearing arrangements, and for the importance of women's work outside the home. Antoinette Brown Blackwell, the first US woman to be ordained a minister, cited Darwin's evidence of male provisioning and female extra-reproductive labour in animals to argue against the Victorian domestic

physician Edward Clarke's 1873 *Sex in Education*, which warned that women's 'enfeebling' menstrual cycles legislated against their participation in higher education, they began to articulate the need for better science relating to women. With lively examples, Hamlin relates how educated middle- and upper-class US women of the era, such as physician Mary Putnam

division of the "eight-hour husband" working outside the home and the "fourteen-hour wife" within it. Feminist intellectual Charlotte Perkins Gilman drew on Darwin's sexual-selection theory to argue that women's economic dependence on men was unnaturally skewing evolution to promote "excessive sexual distinctions". She proposed that economic and reproductive freedom for women would restore female autonomy in choice of mate — which Darwin posited was universal in nature, except in humans — and put human evolutionary progress back on track.

Darwin himself opposed birth control and asserted the natural inferiority of human females. The adult female, he wrote in *The Descent of Man* (1871), is the "intermediate between the child and the man". Nevertheless, appeals to Darwinist ideas by birth-control advocates such as Margaret Sanger led one critic to bemoan in 1917 that "Darwin was the originator of modern feminism".

Feminism in the late nineteenth century was marked by the racial and class politics of the era's reform movements. Blackwell's and Gilman's views that women should work outside the home, for example, depended on the subjugated labour of lower-class minority women to perform household tasks. And Sanger's birth-control politics appealed to contemporary fears of race and class 'suicide'. *From Eve to Evolution* acknowledges this legacy, but does not dwell on it.

For example, Hamlin argues that the anti-biological determinist arguments of white nineteenth-century feminists are more "nuanced and complex" than generally appreciated and were a resource "not just for white women but for everyone". On Gilman, Hamlin asserts that "at least most of the time, Gilman meant the 'human race' when she wrote the 'race' and that racism is not the defining characteristic of most of her writings". Although context is clearly crucial to a careful reading of this complicated intellectual history, it is hard to evaluate such claims without a more detailed treatment of the words and deeds of these feminist writers on the matter of race and class.

Nonetheless, this deeply researched and richly detailed picture of US feminism in the late nineteenth century and early twentieth century is an important contribution to our understanding of the interrelation of gender politics and science. *From Eve to Evolution* firmly corrects the mistaken view that evolutionary biology and feminism are at odds. And it reveals a more diverse dialogue around the science of sexual equality in the era than is generally appreciated. ■

Sarah S. Richardson is a science historian specializing in women, gender and sexuality at Harvard University in Cambridge, Massachusetts. Her latest book is *Sex Itself*. e-mail: srichard@fas.harvard.edu

L-R: TIME LIFE PICTURES/GETTY; BETTMANN/CORBIS/ LIB. OF CONGRESS

INFORMATION TECHNOLOGY

Forgotten prophet of the Internet

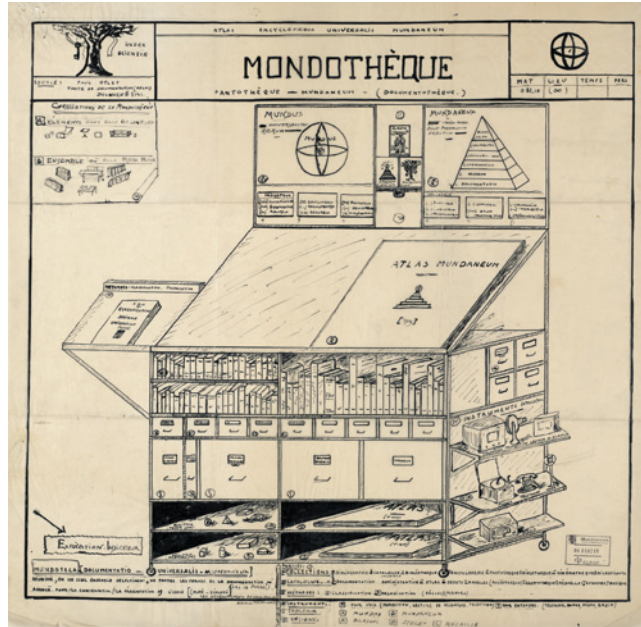
Philip Ball ponders the tale of a librarian who dreamed of networking information.

The Internet is considered a key achievement of the computer age. But as former *New York Times* staffer Alex Wright shows in the meticulously researched *Cataloging the World*, the concept predates digital technology. In the late nineteenth century, Belgian librarian Paul Otlet conceived schemes to collect, store, automatically retrieve and remotely distribute all human knowledge. His ideas have clear analogies with information archiving and networking on the web. Wright makes a persuasive case that Otlet — now largely forgotten — deserves to be ranked among the conceptual inventors of the Internet.

Wright locates Otlet's work in a broader narrative about collation and cataloguing of information. Compendia of knowledge date back at least to Pliny the Elder's *Natural History* (AD 77–79) and the collections of Renaissance scholars such as Swiss naturalist Conrad Gesner, although these were digests of typically uncited sources. Otlet sought to collect everything — newspapers, books, pamphlets, photographs — and to devise a method of categorization that would rival the Dewey decimal system. Wright tells a poignant story of the elderly, perhaps senile, Otlet stacking up jellyfish on a beach and then placing on top an index card bearing the number 59.33: the code for Coelenterata in his Universal Decimal Classification.

Otlet envisaged a 'Mundaneum', a repository of all knowledge. Central to the scheme was the Universal Bibliography, a card index with more than 15 million entries in filing cabinets. Realizing how much space and labour such a system demanded, Otlet advocated the miniaturization of documents (on microfilm) and planned automatic systems to locate information, like steampunk search engines. This knowledge, he thought, might be broadcast to users by radio, and stored in a workstation called a Mondotheque, equipped with microfilm reader, telephone, television and record player.

All this can be correlated with the software and hardware of today. But Wright recognizes that the comparison goes only so far. Otlet's vision was consistent with the social climate of his day: centralized, highly managed and hierarchical. It was quite unlike the distributed, self-organized peer-to-peer



Paul Otlet's Mondotheque workstation.

networks made possible by the personal-computer revolution that was shaped by the counterculture of the 1960s and 1970s.

The real focus of this story is not the antecedents of the Internet. It concerns dreams of a utopian world order, shared by many around the end of the nineteenth century and after the First World War. This was Otlet's grander vision, to which his collecting and cataloguing were merely instrumental. In 1919, with politician Henri La Fontaine — a committed internationalist awarded the 1913 Nobel Peace Prize — Otlet successfully petitioned the Belgian government to fund plans to house the collection in a wing of a grand building in Brussels. He dubbed this space the Palais Mondial. The two men imagined an 'intellectual parliament' for all humanity, in which the organization of knowledge would contribute towards philosopher Auguste Comte's vision of a rationally governed society. In part, their ideas paved the way for the League of Nations and the United Nations — although Otlet was distraught when the Paris Peace Conference of 1919 elected to

establish the league in Geneva, in neutral Switzerland, rather than Brussels. But their objective was much more grandiose, utopian and strange.

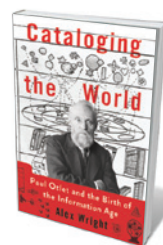
Progressive thinkers such as H. G. Wells (whom Otlet read) desired world government in the interwar period, but Otlet's plans often seemed detached from mundane realities. They veered into mystical notions of transcendence of the human spirit, influenced by theosophy, and Otlet seems to have imagined that learning could be transmitted not only by careful study of documents but by a symbolic visual language in posters and displays. In the late 1920s, he and architect Le Corbusier devised a plan to realize the Mundaneum as a building complex full of sacred symbolism, as much temple as library. Wright overlooks the real heritage of these ideas: Otlet's predecessor here was not Gesner but Italian philosopher Tommaso Campanella, who in 1602

described a "City of the Sun" in which citizens imbibed knowledge from great, complex wall paintings. This aspect of Otlet's dreams looks backwards to Neoplatonism and Gnosticism as much as it looks forward to the information age and the Internet. Perhaps unsurprisingly, politicians remained unconvinced, dooming Otlet to frustration and ultimate failure.

The modern ability to access Isaac Newton's *Principia* online would have delighted Otlet. That so much more network traffic involves cats and pornography would have devastated him. He was devastated enough: the actual Mundaneum never amounted to more than a corner of the building hosting the Palais Mondial, and the government edged him out of there in 1924 to make room for an exhibition on rubber. After losing funding for the cataloguing project in 1934, Otlet clung to a corner of the premises until the Nazis destroyed most of his collection in 1940.

The remainder mouldered for decades in various buildings in Brussels; what survived now sits modestly in the new Mundaneum in Mons, a former garage. But there is another Mundaneum in Brussels: a conference room given that name in Google's European office. It is a fitting tribute, and Wright has offered another. ■

Philip Ball is a writer based in London. His next book, *Invisible*, will be out in August. e-mail: p.ball@btinternet.com



Cataloging the World: Paul Otlet and the Birth of the Information Age
ALEX WRIGHT
Oxford University Press: 2014.

Correspondence

Investigate smog to inform policy

The severe air-pollution episodes that affected Europe this spring (see, for example, go.nature.com/si6mhu and go.nature.com/1b7ygf) highlight the need for effective policy measures and management strategies, given the persistence and potential global reach of such acute events.

Reducing particulate matter, ozone and greenhouse gases is essential to mitigate air pollution as well as climate change, so policies need to be coordinated. For example, climate policy encourages use of fuels such as diesel (because its combustion releases less carbon dioxide per kilometre than petrol) and biomass, which offsets most of its own carbon release as it grows. However, both fuels generate harmful particulates.

For now, the factors that triggered Europe's latest air-pollution episodes and influenced their duration and spread must be investigated. This insight will aid the development of much-needed cross-border policies that are based on sound scientific advice.

Paul S. Monks* *University of Leicester, UK.*
psm7@leicester.ac.uk

**On behalf of 11 co-authors (see go.nature.com/nxe3pb for full list).*

Soil observatory lets researchers dig deep

The UK Soil Observatory (UKSO) launched last month in London. It provides a global web-based platform for the acquisition and dissemination of soil data (www.ukso.org).

Soil security is rising rapidly on government agendas, driven by the United Nations Global Soil Partnership (see *Nature* **492**, 186; 2012). Researchers therefore need to be able to predict how soils will respond to changing climate, vegetation, erosion and pollution (M. W. Schmidt *et al.* *Nature* **478**, 49–56; 2011).

The UKSO is a step change in meeting this challenge.

Expanding on the success of the mySoil crowd-sourcing app (see W. Shelley *et al.* *Nature* **496**, 300; 2013), the UKSO is a community initiative funded by the Natural Environment Research Council. It offers live data streaming and pioneers a route to big data resources. It is comprehensively linked to other sources of soil data, both free and commercial, and to real-time data from the COSMOS national soil-moisture network.

The UKSO interactive map viewer provides access to a huge range of information, from local soil biodiversity to metal concentrations in topsoil.

Data from the UKSO will underpin major government projects, such as the £160-million (US\$268-million) strategy for agricultural technologies (go.nature.com/6oqmxh), and will catalyse associated industrial and commercial enterprises.

Russell Lawley *NERC/British Geological Survey, Environmental Science Centre, Nottingham, UK.*

Bridget A. Emmett, David A.

Robinson *NERC Centre for Ecology and Hydrology, Environment Centre Wales, Bangor, UK.*
davi2@ceh.ac.uk

Strict vaccine quality control in China

We disagree that vaccine quality control is a problem in China (*Nature* **507**, S12–S13; 2014). In 2013, for example, the agency responsible for the country's vaccine quality control — the National Institutes for Food and Drug Control (NIFDC) in Beijing — was designated as a World Health Organization Collaboration Center for Standardization and Evaluation of Biologicals.

Since 2002, China has implemented batch release of vaccines as a key quality-control measure to ensure that they are safe to use. Each batch is

screened by the NIFDC before its release. China produces about 4,000 batches of vaccines annually, corresponding to some 700 million doses.

In 2011, eleven batches of vaccine were rejected because they were of sub-standard quality; three were rejected in 2012 and four in 2013. Together, these represent a total of 1 million doses and correspond to just 0.05% of throughput over three years — a mark of China's stringent vaccine-production processes and of the success of batch-testing practices.

Zhenglun Liang, Qunying Mao, Junzhi Wang *National Institutes for Food and Drug Control, Beijing, China.*
wangjz@nifdc.org.cn

Rethink Campania's toxic-waste scandal

As you point out, it is not easy to prove a correlation between the health issues of people in Italy's Campania region and their exposure to toxic waste (see *Nature* **508**, 431; 2014). The authorities seem to prefer to blame lifestyle factors, but that does not explain why thousands of non-smoking and sober sheep in the area have had to be removed from the food chain.

This waste crisis is a result of certified disposals as well as of illegal dumping, currently being investigated by the European Union project ENTITLE (www.politicaecology.eu); the crisis has been associated with environmental injustice and a disregard for democracy (see G. D'Alisa *et al.* *Ecol. Econ.* **70**, 239–249; 2010). The area has become a political laboratory for testing extreme regulations. It is in a permanent state of emergency, with environmental and civil-rights laws suspended.

Rather than abusing emergency decrees in this way, a different type of laboratory is needed, in which social and natural scientists can

accumulate the facts necessary to address the injustices embodied in the current patterns of contamination and exposure.

Giacomo D'Alisa *Autonomous University of Barcelona, Spain.*

Marco Armiero *Royal Institute of Technology, Stockholm, Sweden.*

Salvatore Paolo De Rosa *Lund University, Sweden.*

giacomo_dalisa@yahoo.it

Internal factors drive Chinese patent surge

China is outpacing many countries in patenting renewable-energy technologies (see J. E. Trancik *Nature* **507**, 300–302; 2014). But this achievement has arguably been inflated by the country's patenting laws.

For example, according to the Espacenet patent database, from 1978 to 2008 China was awarded 21,192 patents on solar photovoltaics. Of these, 99% were granted by China's State Intellectual Property Office. By contrast, the US Patent and Trademark Office granted just 83 patents in this field over the same period (C.-Y. Wu *Scientometrics* **98**, 429–446; 2014).

The overwhelmingly domestic focus of China's patenting activity is a result of its patent-law system, which is designed to promote indigenous innovations. Chinese firms tend to file as many patents as possible for a single equivalent invention, encouraged by extra public subsidies and administrative support (see, for example, A. de la Tour *et al.* *Energy Policy* **39**, 761–770; 2011).

This factor needs to be borne in mind when evaluating Chinese patent figures for the purposes of technological ranking.

Ching-Yan Wu *Fu Jen Catholic University, Taiwan.*

Mei-Chih Hu *National Tsing Hua University, Taiwan.*

John A. Mathews *Macquarie University, Sydney, Australia.*
mchu@mx.nthu.edu.tw

Adolf Seilacher

(1925–2014)

Palaeontologist who pioneered analysis of trace fossils.

Adolf Seilacher used the simplest of methods — careful observation — to transform our understanding of ancient organisms. His interpretations of the enigmatic Ediacara fossils, which date from about 578 million years ago, before the appearance of the major animal phyla during the Cambrian explosion, explained forms found among the earliest large organisms.

Seilacher also showed how trace fossils — those that record biological activity such as the burrowing of marine animals — reveal behavioural traits. He analysed the influences that shape invertebrate morphology and showed how exceptionally preserved fossil assemblages (for which he used the term *Lagerstätten*) are the result of conditions such as low oxygen, rapid burial and the effect of microbial films that seal the sediment surface.

In a career that straddled the Atlantic, Seilacher influenced palaeontology as much with his personality as with his publications. With a commanding yet engaging presence, he had a way of asking questions that caught speakers off guard — an occurrence referred to as being ‘Dolfed’.

Seilacher, who died on 26 April aged 89, was born in 1925 near Stuttgart in Germany. He found his first fossil at the age of 14 and published his first paper, on fossil sharks from local rocks, at 18. He served in the German navy in the last years of the Second World War before entering the University of Tübingen in 1945.

There, palaeontologist Friedrich von Huene taught him to use the camera lucida, an apparatus with a prism and mirror that projects the image of a specimen onto a sheet of paper so that it can be traced. Over the rest of his career, Seilacher made thousands of drawings in this way, both as illustrations for his publications, and as an aid to understanding fossils. He favoured a portable camera lucida that once belonged to von Huene rather than the modern versions that attach to binocular microscopes.

Seilacher remained at Tübingen for his doctorate to study Jurassic and Triassic trace fossils (dating from about 252 million to 145 million years ago). He spent time at the Senckenberg marine station on the Baltic Sea, where he learned ‘actuopalaeontology’ — using the behaviour of living organisms as a guide to interpreting fossils. In 1951, on an expedition to the Salt Range in Pakistan, Seilacher and his PhD adviser



Otto Schindewolf discovered trilobite tracks in early Cambrian rocks, conveying clues about the animals’ lifestyle. In later years, Seilacher’s advice was sought by oil companies exploring sandstones in North Africa and the southwestern United States, because he could use trace fossils to interpret the age and depositional environment of the rocks.

Seilacher spent much of his career at Tübingen, retiring as a professor in 1990. From 1987 to 2009 he spent autumns teaching at Yale University in New Haven, Connecticut. In 1992, he was awarded the Crafoord Prize of the Royal Swedish Academy of Sciences. This enabled him to travel the world, often with his wife, Edith, making replicas of surfaces that preserve an extraordinary range of trace and other fossils and sedimentary structures such as ripples. This collection became his own international touring exhibition, entitled Fossil Art.

Seilacher’s insights fundamentally changed the analysis of the evolution of form. He broke away from the prevailing idea that all morphological features are adaptations for some function. He mined collections, particularly those of the Peabody Museum of Natural History at Yale, for convergent morphologies that provided evidence of other influences. No Seilacher course on invertebrates was complete without his demonstration that a water-filled balloon automatically assumes a form similar to

that of a regular sea urchin, to show that not every shape has adaptive significance.

Seilacher’s ability to interpret morphology from first principles is best demonstrated by his research on the strange fossils of the Ediacaran period, with shapes such as fronds, spindles and discs. He saw in these “an exotic principle of organismic construction” consisting of quilted chambers that maximize the surface available for feeding and respiration. He assigned the Ediacara organisms to an extinct kingdom of life that he called Vendobionta. The relationships of these organisms remain controversial, but many examples investigated by Seilacher on Mistaken Point in Newfoundland, Canada, show a fractal organization that defies easy assignment to any living group.

Among the trace fossils, Seilacher was particularly fascinated by *Paleodictyon*, a regular, often hexagonal, mesh-like structure that is connected to the sediment surface by short vertical tunnels, and occurs in rocks from as far back as the Cambrian. Seilacher interpreted this marine burrow as a ‘farm’ for raising bacteria.

In 1976, modern examples of *Paleodictyon* were discovered in the deep sea by oceanographer Peter Rona. Almost 30 years later, Seilacher and Rona boarded the submersible vessel *Alvin* and retrieved samples from the Mid-Atlantic Ridge at a depth of 3.4 kilometres. Stephen Low’s 2003 film *Volcanoes of the Deep Sea* pictures the two of them slicing soft sediments containing the burrow system — sadly, the animal proved elusive.

An inveterate traveller, Dolf taught and did field work on every continent except Antarctica. He collected not only fossils, but also rugs, ancient cylinder seals and pre-Columbian art, and loved to share stories while enjoying a glass of wine and a cigar. He published works on an extraordinary range of fossils over his 70-year career, notably his 2007 text *Trace Fossil Analysis* (Springer) and his forthcoming *Morphodynamics* (CRC Press).

His iconic drawings and interpretations will remain inspirational for teachers and students for years to come. ■

Derek E. G. Briggs is director of the Peabody Museum of Natural History at Yale University in New Haven, Connecticut, USA. He taught with Dolf Seilacher at Yale from 2003 to 2009.
e-mail: derek.briggs@yale.edu

WOLFGANG GERBER

EPIGENETICS

Keeping one's sex

Progeny of the protist *Paramecium tetraurelia* always retain the parental mating type. This inheritance is revealed to result from an RNA-guided DNA-deletion pathway that protects the genome from foreign DNA sequences. [SEE ARTICLE P.447](#)

DOUGLAS L. CHALKER

Opposites attract. For the unicellular organism *Paramecium tetraurelia*, those opposites are two distinct mating types, called even (E) and odd (O). Sex between E and O cells occurs by conjugation, a reversible cell fusion during which cell partners exchange genetic material before re-separating into progeny cells. Although these offspring all start development with identical, mixed genomes, each cell somehow retains the mating type of its parent: E begets E, and O begets O. The long-standing mystery of this specific inheritance¹ has now been solved by a study on page 447 of this issue, in which Singh *et al.*² report that small RNA molecules communicate mating-type-specific genetic information between parent and progeny cells.

Paramecium are ciliates, members of a phylum of protists that possess both germline and somatic genomes housed in functionally distinct nuclei — the micronucleus and macronucleus, respectively. During conjugation, the micronuclei undergo meiotic division to form gametic nuclei, which are reciprocally exchanged between the partner cells. Fusion of these gametic nuclei generates two genetically identical zygotic genomes within the still-conjugated cells. The cells then separate, the original parental macronuclei are discarded, and new somatic and germline genomes are created from the zygotic genomes in each cell's unique cytoplasm. Although these new genomes represent a combination of the two parental genomes, the mating type of the progeny cells remains the same as before conjugation (Fig. 1).

Formation of the new somatic genome involves extensive genomic remodelling, including the excision of transposable elements — foreign DNA sequences, also called transposons, that can change position in the genome. Also removed are tens of thousands of short sequences (some as small as 26 base pairs), many of which come from within protein-coding regions³. To identify the DNA to eliminate, a pool of small RNA molecules that represents the entire germline genome is produced in the micronuclei shortly after conjugating cells pair. These small RNAs are initially transported to the parental macronucleus

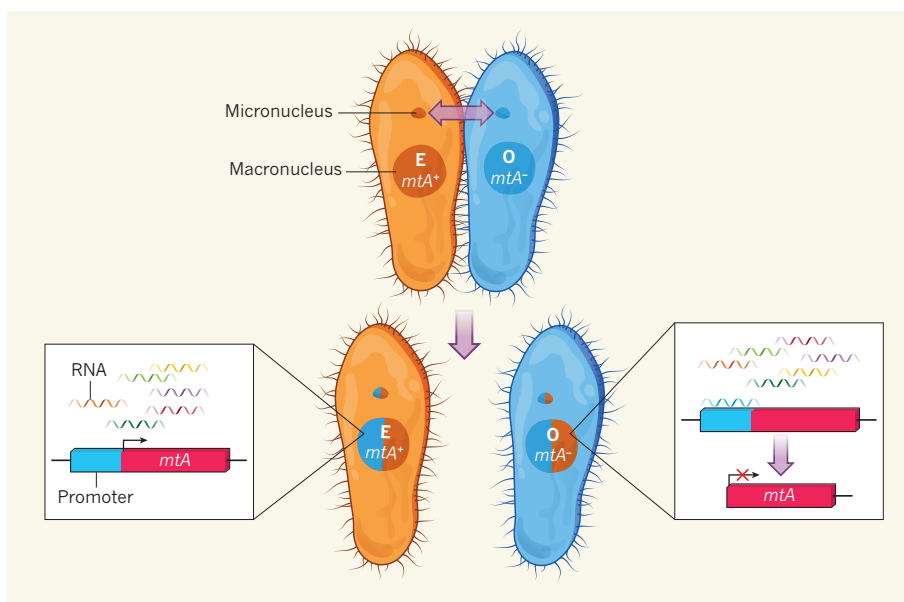


Figure 1 | Epigenetic inheritance of mating type. The unicellular organism *Paramecium tetraurelia* exists in two mating types: E cells, which express the cell-surface protein mtA, and O cells, which do not. The cells contain a somatic genome housed in a macronucleus and a germline genome in a micronucleus. During sexual reproduction, parental E and O cells conjugate and exchange germline genomes, which give rise to new, mixed (bi-coloured) germline and somatic genomes after the cells re-separate. During this process, a pool of RNA molecules is generated that targets foreign regions of the new genomes for deletion, by comparing them with the parental somatic genome. Singh *et al.*² show that this RNA-guided DNA-deletion pathway results in mating-type inheritance in this organism: the promoter region of the gene *mtA* is always deleted in the progeny of O cells, but never in the progeny of E cells, because it is present only in the E cells' parental genome.

before it is discarded, and any RNAs for which there are matching sequences in the parental somatic genome are removed from the pool. (The actual comparison probably occurs between meiotic small RNAs and somatic transcripts expressed from the parental genome⁴.)

The remaining small RNAs — which represent sequences present in the germline genome but absent from the parental somatic genome, and which thus primarily correspond to foreign DNA — are bound by molecules called Piwi-related proteins and carried to the developing somatic genome, where any sequences for which there are homologous small RNAs are targeted for excision. The end result is a reorganized genome that is modelled on that of the parent. This RNA-mediated DNA-elimination pathway is the ciliates' evolutionary equivalent of the Piwi-interacting (pi)RNA pathway that silences transposons in the germline cells

of multicellular organisms⁵. Singh *et al.* show that *Paramecium* has adapted this pathway to regulate its sex determination.

The authors find that E cells are attractive to O cells because the E cells express a cell-surface protein called mtA. This protein is expressed only in E cells because the promoter sequence of the *mtA* gene behaves like a transposon-derived sequence in O-cell somatic nuclei and so is targeted for deletion (Fig. 1). During development, E-cell macronuclei do not delete the promoter because that sequence exists in the somatic genome of the E parent. Thus, the presence or absence of the *mtA* promoter is communicated between parent and progeny genomes by homologous RNAs, which effectively block or promote DNA deletion, respectively, thereby ensuring that mating type is propagated to the next generation. This process can be described as an epigenetic

phenomenon, because different patterns of gene activity are inherited even though the two cells start with the same DNA sequence.

Although the DNA-elimination pathway is known only in ciliates, RNA-mediated gene silencing is not, and several studies have indicated that small RNAs commonly function to mediate transgenerational inheritance. For example, in *Drosophila* fruit flies, piRNAs produced maternally from regions near the ends of chromosomes can silence transposons or other homologous sequences that reside elsewhere in the genome⁶. In *Caenorhabditis elegans* worms, three small-RNA pathways cooperate to repress experimentally introduced genes that were not present or expressed in the previous generation, thereby serving to identify new foreign sequences in the genome^{7–9}. In these cases, the expressed genome content of the current generation is communicated to the next by comparison of small RNAs with available mRNA transcripts, which are produced from protein-coding regions of the genome. In *Paramecium*, however, Singh *et al.* show that both coding and non-coding sequences can be compared, because full-genome complements of both germline small RNAs and somatic transcripts are generated during the cells' development².

The piRNA pathways have largely been regarded as a means of distinguishing self from non-self and thereby silencing foreign genome invaders (such as transposable elements). But Singh and colleagues have shown that RNA-mediated genome-defence pathways can be exploited to evolve modes of epigenetic regulation. The co-option of this pathway to regulate genes may be more frequent than this one case reveals. When the authors examined related *Paramecium* species that also display epigenetic inheritance of mating types, they found that *Paramecium septaurelia* does not differentially excise the *mtA* promoter in different mating types but, instead, O cells delete a block of sequence from another gene, *mtB*, which encodes a transcription-regulating protein required for *mtA* expression. The excision creates a non-functional version of this gene and so *mtA* is not expressed. E cells of this species, however, retain this sequence and express a functional *mtB* protein and therefore also *mtA*.

Thus, even for this one trait, the DNA-excision pathway has been adapted twice to epigenetically regulate two different genes. It has taken nearly 75 years to uncover how mating type is inherited during sexual reproduction in *Paramecium*. We can hope that this discovery will expedite the identification of other examples of cell-intrinsic gene regulation mediated by homologous RNAs. ■

Douglas L. Chalker is in the Department of Biology, Washington University in St. Louis, St. Louis, Missouri 63130, USA.
e-mail: dchalker@wustl.edu

1. Sonneborn, T. M. *Adv. Genet.* **1**, 263–358 (1947).
2. Singh, D. P. *et al. Nature* **509**, 447–452 (2014).
3. Arnaiz, O. *et al. PLoS Genet.* **8**, e1002984 (2012).
4. Lepère, G., Bétermier, M., Meyer, E. & Duhaucourt, S. *Genes Dev.* **22**, 1501–1512 (2008).
5. Malone, C. D. & Hannon, G. J. *Cell* **136**, 656–668 (2009).

6. Josse, T. *et al. PLoS Genet.* **3**, 1633–1643 (2007).
7. Ashe, A. *et al. Cell* **150**, 88–99 (2012).
8. Lee, H.-C. *et al. Cell* **150**, 78–87 (2012).
9. Shirayama, M. *et al. Cell* **150**, 65–77 (2012).

This article was published online on 7 May 2014.

ASTROPHYSICS

Windy stars that go with a bang

The probable signature of an ageing massive star's stellar wind has been detected shortly after the star underwent a supernova explosion. The finding suggests that such windy stars can have bright, observable deaths. [SEE LETTER P.471](#)

JOHN J. ELDRIDGE

The most direct indication yet found that nearly all massive stars end their lives in supernovae is described by Gal-Yam *et al.*¹ on page 471 of this issue. The authors report the detection of emission lines in a spectrum taken just 15.5 hours after a star exploded as a supernova, uncovering the nature of the progenitor star. The spectral features closely resemble those expected for Wolf–Rayet stars — ageing massive stars that shed material rapidly in strong stellar winds (Fig. 1). Until this event, there was growing evidence that such

stars were likely to have dim or unobservable deaths.

Supernovae are classified by their observed characteristics. For example, those with hydrogen lines in their spectra fall into the type II category, and those without them are classified as type I. This spectral signature provides information about the composition of the star that exploded, and the nature of the explosion. The presence of silicon lines indicates a type Ia supernova produced by a dense white dwarf star that underwent a thermonuclear explosion.

The more common events, known as core-



Figure 1 | An ageing massive star. Wolf–Rayet stars such as the one shown here (central bright object) are evolved, massive stars that continually lose gas in dense stellar winds (orange and yellow glow). Gal-Yam *et al.*¹ identified Wolf–Rayet-like wind signatures in a spectrum taken 15.5 hours after a star exploded as a supernova, before the expanding supernova ejecta could sweep up the wind.

collapse supernovae, include types II, Ib and Ic. They do not have silicon lines in their spectra, and result from the collapse of the core of a massive star. The temperatures at the centre of such massive stars are high enough for nuclear fusion in their cores to progress from hydrogen all the way to a mixture of iron, cobalt and nickel. But fusion of these elements is not energetically favourable, and so the core collapses to a neutron star — or to a black hole in the case of the most massive stars. The gravitational energy released by the collapse is transferred to the material surrounding the core, causing the star to explode.

The only events for which the progenitor stars are definitely known are the type II supernovae²: they are mainly red supergiant stars, which are rich in hydrogen and have cool surface temperatures. Many of these stars have been directly identified in images fortuitously taken before the stars exploded. Their masses fall into a narrow range, from 8 to 20 solar masses^{2–5}. There are also a few examples of rare and extreme supernovae that may have had as progenitors luminous blue variable stars, which are thought to be around 100 solar masses. But this is difficult to confirm, because no spectra of such progenitor stars have been obtained⁶.

According to theoretical models of stellar evolution, Wolf–Rayet stars, which are greater than 20 solar masses, should end their lives in core-collapse supernovae and have strong stellar winds throughout their lifetime. The latter property means that these stars will lose all, or nearly all, of their hydrogen by the time they explode. The low hydrogen abundances make the stars extremely luminous and hot, which in turn causes the dense stellar wind to be driven from the star's surface at a speed of at least 1,000 kilometres per second. The signature of the wind is imprinted on the spectrum of the star⁷ in the form of broad emission lines, with the widths of the lines revealing the wind speed.

However, until Gal-Yam and colleagues' study, there was no direct evidence that Wolf–Rayet stars do explode in supernovae. Statistical considerations indicate that pre-explosion images of a sample of 12 nearby supernovae that are poor in or devoid of hydrogen should have revealed at least one, but none has been found⁸. Also, last year's detection of a supernova progenitor that seemed to have properties consistent with those of Wolf–Rayet stars is now thought not to be associated with such stars⁹. Their fate has therefore remained uncertain, and it has been suggested that they end their lives in weak or invisible explosions¹⁰.

By identifying Wolf–Rayet-like wind signatures in a spectrum of a supernova obtained less than one day after the star exploded, Gal-Yam and colleagues show that at least one Wolf–Rayet star may have ended its life in a visible supernova. Key to the authors' finding was both the ability to obtain a spectrum so quickly after the explosion, before the stellar

wind was swept up by the explosion, and the fact that Wolf–Rayet stars have winds that extend great distances from the star. Detecting Wolf–Rayet-like spectral features in an early-supernova spectrum is by far the most direct evidence that these stars have bright and observable deaths. The only way to strengthen this conclusion would be to obtain a spectrum of the star before it exploded.

But Gal-Yam and colleagues' results go beyond simply identifying the type of star that exploded. The authors have also been able to estimate the speed and density of the stellar wind. If similar observations are made for other ageing windy stars, then we will begin to understand the range of possible stellar-wind properties that such stars could have shortly before they die, and to learn whether these death throes are very different and special, or only as expected. Either way, they will provide important clues about how the most massive stars end their lives. ■

PHYSIOLOGY

Double function at the blood–brain barrier

Two aspects of the blood–brain barrier — the transport of lipids to the brain and the transport of molecules across cells lining blood vessels — have been shown to be regulated by the same protein, Mfsd2a. SEE LETTERS P.503 & P.507

CHRISTER BETSHOLTZ

The blood–brain barrier (BBB) is a double-edged sword. On the one hand, this cellular interface helps to maintain a constant, optimal environment for neuronal function through a combination of barriers and selective transport systems that regulate the passage of wanted and unwanted molecules. But on the other hand, it presents a formidable challenge to medicine because it stops most drugs from passing from the bloodstream to the brain. Two papers in this issue add considerably to our rudimentary understanding of the BBB: Nguyen *et al.*¹ (page 503) unravel how an essential omega-3 fatty acid is transported across it, and Ben-Zvi *et al.*² (page 507) identify a mechanism involved in suppressing the vesicle-mediated transfer of blood-plasma constituents to the brain.

Astonishingly, these two seemingly unrelated processes depend on the same gene, *Mfsd2a*, which encodes a transmembrane protein that is specific to the endothelial cells that line blood vessels in the brain. Two birds killed with one stone, it seems. But the discoveries also illustrate the contradictory properties of the BBB: Mfsd2a transports lipids that are

John J. Eldridge is in the Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand.
e-mail: j.eldridge@auckland.ac.nz

- Gal-Yam, A. *et al.* *Nature* **509**, 471–474 (2014).
- Smartt, S. J. *Annu. Rev. Astron. Astrophys.* **47**, 63–106 (2009).
- Li, W. *et al.* *Astrophys. J.* **661**, 1013–1024 (2007).
- Maund, J. R., Smartt, S. J., Kudritzki, R. P., Podsiadlowski, P. & Gilmore, G. F. *Nature* **427**, 129–131 (2004).
- Benvenuto, O. G., Bersten, M. C. & Nomoto, K. *Astrophys. J.* **762**, 74 (2013).
- Gal-Yam, A. & Leonard, D. C. *Nature* **458**, 865–867 (2009).
- Crowther, P. A. *Annu. Rev. Astron. Astrophys.* **45**, 177–219 (2007).
- Eldridge, J. J., Fraser, M., Smartt, S. J., Maund, J. R. & Crockett, R. M. *Mon. Not. R. Astron. Soc.* **436**, 774–795 (2013).
- Fremming, C. *et al.* Preprint at <http://arxiv.org/abs/1403.6708> (2014).
- Kochanek, C. S. *et al.* *Astrophys. J.* **684**, 1336–1342 (2008).

essential for brain growth and function, while at the same time suppressing a transport route across the BBB that might be ideal for the delivery of complex drugs, such as antibodies, into the brain.

Mfsd2a is a member of the major facilitator superfamily (MFS) of secondary-active transporters³ — proteins that use the electrochemical potential of solutes to shuttle specific substrates across lipid membranes. However, many MFS transporters do not yet have described substrates and functions⁴. Previous work has shown that *Mfsd2a* expression is induced by factors in the liver during fasting⁵, and the protein has also been implicated in antibiotic transport⁶ and in cell fusion in the human placenta^{7,8}. However, Mfsd2a is strongly evolutionarily conserved, which seems inconsistent with a human-specific role in the placenta, and instead suggests that its primary function lies elsewhere. Moreover, Mfsd2a is found at only low levels in the liver, but it is highly expressed throughout the brains of several species, and mice lacking Mfsd2a have normal liver metabolism but develop neurological disorders⁹. Now, both Nguyen *et al.* and Ben-Zvi *et al.* determine that Mfsd2a is specifically and constitutively expressed by

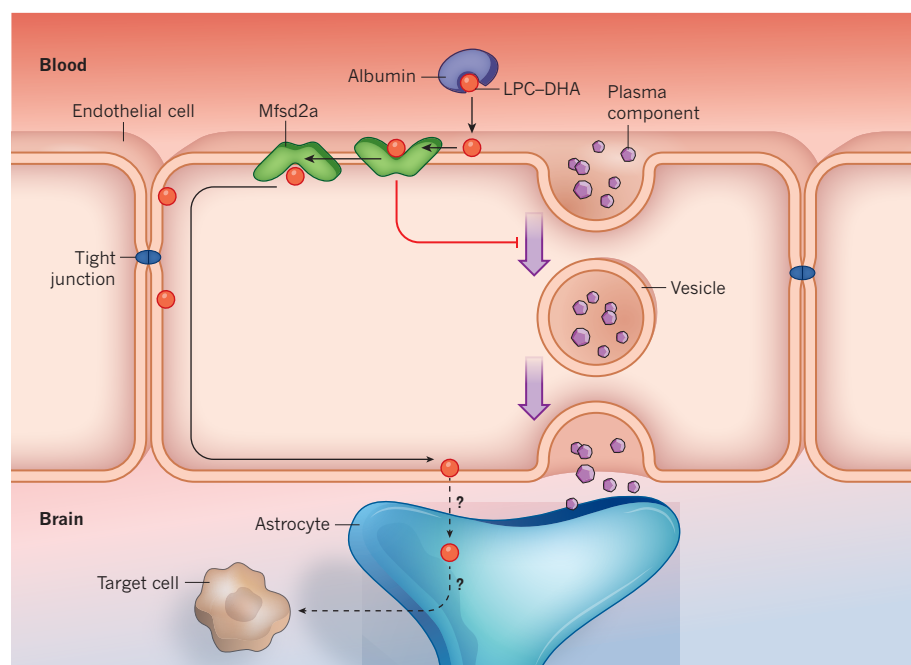


Figure 1 | Dual role for Mfsd2a. Lysophosphatidylcholine–docosahexaenoic acid (LPC–DHA) is bound to the protein albumin during transport in the blood. When it reaches the blood–brain barrier, LPC–DHA detaches from albumin and is adsorbed to the outer lipid leaflet of the cell membrane of endothelial cells. Nguyen *et al.*¹ show that Mfsd2a binds LPC–DHA and transfers it to the inner lipid leaflet, potentially allowing it to bypass the tight junctions between endothelial cells and reach the brain-facing side of the cell. The targets of DHA in the brain are unknown, as is the method by which it passes through the ‘end-feet’ structures of astrocyte cells, which completely cover the brain’s blood vessels. Ben-Zvi *et al.*² also show that Mfsd2a suppresses transcytosis in endothelial cells, a transport mechanism by which components of the blood plasma, including proteins, are transferred across the cell in vesicles.

brain endothelial cells, suggesting that it has a role in the BBB.

Nguyen *et al.* extended the anatomical and behavioural characterization of mice lacking *Mfsd2a*, finding that these animals have small brains and a range of motor and cognitive defects, and have reduced numbers of certain neuron types. Noting that these defects were reminiscent of omega-3 fatty-acid deficiency, the authors compared the lipid composition of brains from *Mfsd2a*-knockout and control mice and found reduced levels of a certain omega-3 fatty acid, docosahexaenoic acid (DHA), in the mutant animals. *In vitro* analysis revealed that Mfsd2a transports DHA only when this fatty acid is attached to the lipid lysophosphatidylcholine (LPC). *In vivo* experiments confirmed that the brain’s uptake of DHA occurs mainly through Mfsd2a-dependent transport of LPC–DHA (Fig. 1). Although the importance of DHA for normal brain growth and function was already known¹⁰, the role of LPC as its carrier across the BBB, by means of Mfsd2a, represents a breakthrough in our understanding of how essential fatty acids enter the brain.

Ben-Zvi and colleagues approached Mfsd2a from a different angle. They searched for gene transcripts whose expression by vascular cells correlated with the development of the BBB. Previous work had indicated that, in mammals, the BBB forms, at least in part,

before birth, but it was not clear exactly when this occurs and whether there are regional differences. To observe the development of the BBB, the authors injected fluorescent tracer dyes into the livers of fetal mice, from which the dyes entered the systemic circulation. The authors were then able to map the maturation of the embryonic BBB with high spatio-temporal precision and thus to select a suitable development stage for transcript profiling. They found that Mfsd2a was prominently overexpressed in the brain-cortex endothelium compared with the lung vascular endothelium at embryonic day 13.5, and that the BBB in *Mfsd2a*-knockout mice was impaired from embryonic day 15.5 to adulthood.

The authors also found that, compared to control mice, the brain endothelial cells of *Mfsd2a*-knockout mice displayed increased transcytosis — the transfer of molecules from one side of a cell to another in membrane-bound vesicles (Fig. 1). Increased transcytosis has also been reported for mice with a reduced density of pericytes (cells that wrap around the endothelial cells of small blood vessels) in the brain^{11,12}, which led Ben-Zvi *et al.* to the finding that the expression of Mfsd2a in brain endothelial cells depends on the presence of pericytes. This transcytotic route is particularly attractive from a drug-delivery perspective because there is no evidence for restrictions on the molecular mass or physico-chemical properties of the

cargo transported by this process¹¹. Targeting pericytes as a way of opening the BBB for drug passage would be illogical (because these cells lie on the brain side rather than the blood side of the barrier), but targeting Mfsd2a in the endothelial-cell luminal membrane (which contacts the blood) could be feasible.

These two studies provide the first molecular handle on lipid and membrane-vesicle transport across the brain endothelium, but there are still several details to be explored. How does Mfsd2a regulate endothelial transcytosis? Is it a direct mechanism, or does it act indirectly through deficient lipid transport to the brain? Other causes of lipid deficiency, such as a dietary lack of DHA, might shed light on this question.

Regarding the transport of DHA across the BBB, only the first step has been elucidated so far, and it will be interesting to explore how the lipid (or lipids) transported by Mfsd2a completes its passage across the multicellular BBB and into the brain. Nguyen and colleagues’ findings suggest that Mfsd2a translocates LPC–DHA from the outer leaflet of the endothelial-cell luminal membrane to the inner leaflet. Lipids in the inner leaflet, but not the outer leaflet^{13,14}, can bypass the tight junctions between cells, which may allow diffusion of LPC–DHA from the luminal to the abluminal membrane, which faces the brain, but how further transport of DHA in the brain occurs is unclear. The functions of DHA in the brain are also not known, although possibilities include structural roles in membranes or signalling roles in regulating cell behaviour. Further studies of *Mfsd2a*-knockout mice should allow these and other questions to be addressed. ■

Christer Betsholtz is in the Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 75185, Sweden, and the Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. e-mail: christer.betsholtz@igp.uu.se

1. Nguyen, L. N. *et al.* *Nature* **509**, 503–506 (2014).
2. Ben-Zvi, A. *et al.* *Nature* **509**, 507–511 (2014).
3. Shi, Y. *Annu. Rev. Biophys.* **42**, 51–72 (2013).
4. Law, C. J., Maloney, P. C. & Wang, D.-N. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).
5. Kadereit, B. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 94–99 (2008).
6. Reiling, J. H. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 11756–11765 (2011).
7. Esnault, C. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 17532–17537 (2008).
8. Toufaily, C. *et al.* *Placenta* **34**, 85–88 (2013).
9. Berger, J. H., Charron, M. J. & Silver, D. L. *PLoS ONE* **7**, e50629 (2012).
10. Kidd, P. M. *Altern. Med. Rev.* **12**, 207–227 (2007).
11. Armulik, A. *et al.* *Nature* **468**, 557–561 (2010).
12. Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. *Nature* **468**, 562–566 (2010).
13. Dragsten, P. R., Blumenthal, R. & Handler, J. S. *Nature* **294**, 718–722 (1981).
14. van Meer, G. & Simons, K. *EMBO J.* **5**, 1455–1464 (1986).

This article was published online on 14 May 2014.

Selectivity from flexibility

Porous materials called metal–organic frameworks hold promise for many applications, including molecular separations. One such material has been discovered that shape-shifts to amplify its selectivity for a target molecule.

RYOTARO MATSUDA

One of the most common preconceptions about crystals is that they are rigid. However, this is not necessarily the case for crystals of metal–organic frameworks (MOFs) — crystalline materials in which metal ions are linked by various organic molecules¹. Some MOF crystals can flexibly change their crystalline structures when they trap or release molecules. Writing in *Angewandte Chemie*, Warren *et al.*² describe a flexible MOF that reconstructs its molecular framework to selectively recognize different isomers of xylene, an industrially useful compound. The finding provides much-needed insight into how flexible MOFs should be designed to maximize their selectivity for target molecules.

MOF crystals contain thousands of pores, in most cases of less than 2 nanometres in diameter. These nanoporous materials are therefore being extensively studied for potential industrial applications such as gas storage,

molecular separation and catalysis, along with other porous materials, for example zeolites and activated carbons. At the dawn of MOF chemistry in the 1990s, the primary target was to fabricate rigid, robust and hard crystalline MOFs to achieve highly porous, stable networks³. However, in the 2000s, flexible MOFs emerged that could undergo reversible structural changes without collapsing.

The flexible nature of these MOFs mainly derives from the ‘coordination’ bonds that form between the metal ions and the organic linkers, and the large volume of free space within MOFs. Because coordination bonds are rather weak, bond-breaking and re-formation can occur easily in MOF crystals^{4,5}, and this may, in turn, allow dynamic rearrangement of the bonding geometry. And because the infinite molecular network of MOFs is formed entirely by coordination bonds, such shifts enable the structural transformation of an entire crystal.

The dynamic motion of the organic molecules in MOF crystals also has a significant role

in conferring structural flexibility. Molecules that form part of the porous skeleton have less contact with other components of the MOF than molecules found in most other solids, allowing highly flexible rotational and vibrational movements that enable modulation or tuning of the pore structure.

Initially, flexible transformations were observed when solvent was removed from or adsorbed to MOFs^{6–8}. Subsequently, the phenomenon was also observed to occur when MOFs trapped gas molecules in a dry atmosphere^{9,10}, which ruled out the possibility that the changes were simply the result of the framework recrystallizing. Investigations of the molecular-trapping mechanism in flexible MOFs have advanced considerably as a result of direct observations^{11,12} of framework changes, and of trapped molecules in the nanopores, using X-ray structural analysis.

In general, these flexible systems open their pores to accommodate guest molecules, then return to a closed state once the guests have been discharged. However, a complication for molecular-separation applications is that, once the pores have been opened by the target molecules in a mixture, other unwanted components of the mixture can also diffuse into the pores, provided that they can fit into the available void space within the open-pore form of the framework. Crucially, this can diminish the selectivity of the MOF towards the desired target, and thus reduce its overall utility for separation¹³. It is therefore necessary to establish guidelines for designing flexible MOFs with superior separation abilities to those of existing ones¹⁴.

Enter Warren *et al.*, who describe a MOF that takes up the *para*-isomer of xylene highly selectively over the *meta*-isomer, and that separates *para*-xylene from its other isomers more effectively than materials currently in use (inorganic zeolites). Xylene is obtained as a mixture of isomers from a catalytic process, but *para*-xylene is the most useful, because it is a feedstock for the production of terephthalic acid — which, in turn, is needed to manufacture polyethylene terephthalate, a polymer commonly used to make plastic bottles. Being able to obtain pure *para*-xylene is therefore important; this is not easy, however, because its size and physical properties (such as its boiling point) are similar to those of the other isomers.

Not only do Warren and co-workers report a high-performance separation of *para*- and *meta*-xylene, but they also provide clear crystallographic evidence of the mechanism involved, thus supplying crucial information about how to design flexible MOFs. Their MOF has two types of one-dimensional channel. Notably, the channels can undergo structural relaxation involving geometric changes around the metal (cerium) ions and rotation of a benzene ring in the organic linkers.

When the authors modelled the guest-free framework computationally as a rigid

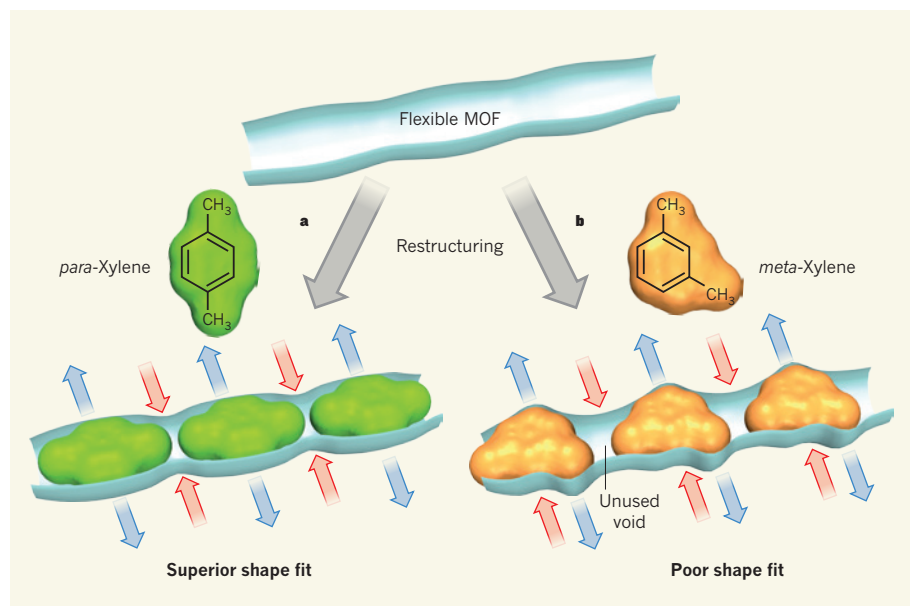


Figure 1 | Pore remodelling enhances molecular recognition. Warren *et al.*² report a metal–organic framework (MOF) containing one-dimensional channels that change shape to bind the *para*- and *meta*-isomers of xylene. **a**, Channel restructuring (red and blue arrows) allows an excellent fit to *para*-xylene molecules. **b**, Such an efficient fit is not possible for *meta*-xylene, and generates unused voids between guest molecules. This difference greatly enhances the MOF's selectivity for *para*-xylene.

structure, the simulation predicted that *meta*-xylene should be excluded from the pores. But, in fact, reconstruction of the framework causes channel relaxation and can allow *meta*-xylene to be accommodated, as proved by X-ray crystallographic analysis. So why does the MOF absorb *para*-xylene selectively?

Warren *et al.* performed further detailed structural studies and found that channel restructuring that results from *para*-xylene accommodation occurs synergistically with positive feedback, which enhances the fitting of the channels to the molecules through localized expansion and contraction (Fig. 1a). Conversely, accommodation of *meta*-xylene requires a large distortion of the channels and occurs with negative feedback, which generates unfavourable unused void space and disordered guest molecules (Fig. 1b). The researchers suggest that the favourable and unfavourable channel restructuring for *para*-xylene and *meta*-xylene, respectively, amplifies the MOF's selectivity in separations of the two isomers.

Although Warren and colleagues clearly

show how the MOF accommodates pure *para*-xylene or *meta*-xylene, it is not clear how the channels trap a mixture of both substrates during separation. This should be studied in future work. Furthermore, the authors studied their system mostly at equilibrium states, but kinetic studies of the separation will provide more practical information for the design of flexible MOFs for different applications.

The conformation of biomolecules such as proteins can be flexibly transformed to enhance their ability to recognize a specific substrate. The new findings prove that this concept also applies to flexible MOFs, and open up a field of molecular recognition and separation by crystalline porous solids. Once guidelines have been established for how to design frameworks to enhance selectivity for a wide range of target molecules, flexible MOFs will be highly promising materials for molecular separation in industry. ■

Ryotaro Matsuda is at the Institute for Integrated Cell–Material Sciences,

Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8530, Japan.

e-mail: rmatsuda@icems.kyoto-u.ac.jp

1. Kitagawa, S., Kitaura, R. & Noro, S. *Angew. Chem. Int. Edn* **43**, 2334–2375 (2004).
2. Warren, J. E. *et al.* *Angew. Chem. Int. Edn* **53**, 4592–4596 (2014).
3. Yaghi, O. M. *et al.* *Nature* **423**, 705–714 (2003).
4. Kitagawa, S. & Matsuda, R. *Coord. Chem. Rev.* **251**, 2490–2509 (2007).
5. Horike, S., Shimomura, S. & Kitagawa, S. *Nature Chem.* **1**, 695–704 (2009).
6. Matsuda, R. *et al.* *J. Am. Chem. Soc.* **126**, 14063–14070 (2004).
7. Serre, C. *et al.* *J. Am. Chem. Soc.* **124**, 13519–13526 (2002).
8. Cussen, E. J., Claridge, J. B., Rosseinsky, M. J. & Kepert, C. J. *J. Am. Chem. Soc.* **124**, 9574–9581 (2002).
9. Kitaura, R., Seki, K., Akiyama, G. & Kitagawa, S. *Angew. Chem. Int. Edn* **42**, 428–431 (2003).
10. Li, D. & Kaneko, K. *Chem. Phys. Lett.* **335**, 50–56 (2001).
11. Matsuda, R. *et al.* *Nature* **436**, 238–241 (2005).
12. Halder, G. J. & Kepert, C. J. *J. Am. Chem. Soc.* **127**, 7891–7900 (2005).
13. Li, J.-R., Sculley, J. & Zhou, H.-C. *Chem. Rev.* **112**, 869–932 (2012).
14. Sato, H. *et al.* *Science* **343**, 167–170 (2014).

CANCER

Darwinian tumour suppression

Competition for access to a survival factor has been found to explain why incoming cells from the bone marrow replace resident cells in the thymus. Reducing this competition can cause tumours to form. SEE ARTICLE P.465

EDUARDO MORENO

The idea that cells in the body compete in a Darwinian manner — that is, according to the laws of natural selection — has been postulated several times. First, in 1881, zoologist Wilhelm Roux anticipated that cells and ‘parts’ of an organism must fight with each other to form the body during embryonic development. A few years later, neuroscientist Santiago Ramón y Cajal predicted that neurons undergo a competitive struggle for space and nutrition. This suggestion gained support with neurologist Rita Levi-Montalcini’s discovery of growth factors, and their integration into the ‘neurotrophic theory’, which proposed that neurons compete for limiting amounts of survival-promoting factors, a process that potentially eliminates unfit cells. On page 465 of this issue, Martins *et al.*¹ describe another example in which Darwinian competition between cells occurs naturally — in this case, serving as a potent tumour-suppressor mechanism in the thymus.

The thymus is a specialized organ of the immune system. Located in the thoracic cavity,

its function is to host a type of immature white blood cell, the T cell, and to develop such cells into mature T cells that are capable of recognizing harmful foreign substances.

The normal function of the thymus depends on a continuous supply of cells from the bone marrow². When these bone-marrow-derived progenitors reach the thymus, they replace thymus-resident progenitors (Fig. 1a).

Martins and colleagues were inspired by previous research in *Drosophila* fruit flies showing that slowly proliferating cells are recognized and eliminated by cells proliferating at a normal rate, through a mechanism proposed to involve competition for extracellular factors³. In a distinctive case of the ‘random walks’ by which science moves forward, those researchers were themselves motivated by the trophic theories⁴ discussed above: that Darwinian-like competition among cells for factors required for survival leads to elimination of a fraction of the cell population.

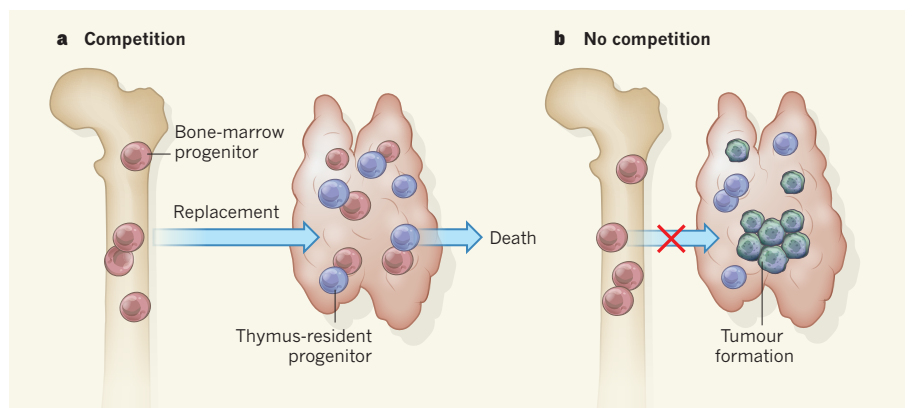


Figure 1 | Cellular competition prevents tumour formation. **a**, T cells develop in the thymus from progenitor cells. The thymus is continuously supplied with progenitor cells from the bone marrow, which replace thymus-resident progenitor cells. Martins *et al.*¹ show that competition between these cells for the survival factor IL-7 (not shown) leads to the death of the thymus-resident progenitors. **b**, The authors also show that an absence of competition, when the supply of bone-marrow progenitors is blocked, leads to uncontrolled self-renewal of thymus-resident progenitors and T-cell development, such that tumours form.

Martins *et al.* resolved to analyse whether, in mice, the normal replacement of thymus-resident progenitors by bone-marrow-derived colonizing cells shows the hallmarks of this type of competition. They found that competition for the blood-cell survival factor interleukin-7 (IL-7) could explain this replacement. IL-7 can activate the expression of an intracellular pro-survival protein, Bcl2 (ref. 5). The authors propose that IL-7 availability is limited for thymus-resident progenitors and that, in the presence of bone-marrow-derived progenitors, this leads to lower levels of Bcl2 in thymus-resident cells and, therefore, their death.

The next (and perhaps the most intriguing) question the researchers asked was: what would be the consequences of a lack of competition in this organ? In other words, what would happen in the absence of incoming bone-marrow progenitors? It was known that in the absence of incoming cells, the thymus-resident progenitors can self-renew and produce T cells⁶, compensating for this loss. But the surprise came when Martins and colleagues observed that the lack of competition from colonizing bone-marrow cells resulted in the formation of tumours (Fig. 1b), owing to the genetic transformation of the thymus-resident progenitors. Interestingly, the transformation of the thymus-resident progenitors produced a tumour type that resembled human T-cell acute lymphoblastic leukaemia in many aspects, including the type of genomic changes, the cells' gene-transcription profiles and the presence of activating mutations in the gene *Notch1*. The results suggest that cell competition is required to periodically replace thymus-resident progenitors with fresh bone-marrow-derived progenitors, and that if this process is interrupted the thymic cells become cancerous.

Thus, this study describes an exciting Darwinian mechanism that functions to prevent cancer in the thymus. But there are several aspects to this process that warrant further investigation. For example, it seems from the authors' findings that cells in the thymus have a higher propensity to develop tumours than cells in other organs, but why might this be the case? Is the environment of the thymus cancer-promoting — do progenitors suffer cellular or genetic insults while performing their tasks in the thymus? Or do cells moving from the bone marrow to the thymus come with a pre-programmed 'expiry date' and become malignant after this?

It is also apparent from work in *Drosophila* that trophic theories are inadequate to explain cell competition. In flies, the amount of survival factor for which cells compete does not need to be limiting, because there is a fascinating mechanism — based on 'fitness fingerprints' displayed on cell membranes — that allows cells to compare their fitness directly^{7,8}. If similar systems enable mammalian cells to exchange fitness information, thymic progenitors might be able to detect and eliminate

less-efficient cells even if there is no scarcity of the molecules they compete for. One testable prediction of this model is that progenitors with impaired IL-7 signalling will modify their fitness fingerprints accordingly, and actively reveal their fitness to neighbouring cells.

What are the consequences of Martins and colleagues' discovery for cancer treatment? T-cell acute lymphoblastic leukaemia is an aggressive cancer that frequently shows more-pronounced resistance to chemotherapy than does the more common form of blood cancer, B-cell lymphoma. Targeting the genes involved in cell competition might provide an avenue for treating and diagnosing T-cell leukaemia.

This work may also have immediate consequences for gene-therapy treatments in humans, because the scenario described by Martins *et al.* in mice also occurs in patients receiving gene therapy for severe combined immunodeficiency — in a subset of patients, the genetically modified precursor cells that are administered generate T cells in the absence of bone-marrow-derived precursors. It now seems that this lack of competition may predispose those patients to tumour

formation. Despite these open questions and the various directions in which this research may develop, it is clear that exciting days lie ahead for the study of cell competition in relation to cancer^{9,10}. ■

Eduardo Moreno is at the Institute of Cell Biology, IZB, University of Bern, Bern CH-3012, Switzerland.
e-mail: emoreno@izb.unibe.ch

1. Martins, V. C. *et al.* *Nature* **509**, 465–470 (2014).
2. Frey, J. R., Ernst, B., Surh, C. D. & Sprent, J. *J. Exp. Med.* **175**, 1067–1071 (1992).
3. Moreno, E., Basler, K. & Morata, G. *Nature* **416**, 755–759 (2002).
4. Raff, M. C. *Nature* **356**, 397–400 (1992).
5. Jiang, Q. *et al.* *Mol. Cell. Biol.* **24**, 6501–6513 (2004).
6. Martins, V. C. *et al.* *J. Exp. Med.* **209**, 1409–1417 (2012).
7. Rhiner, C. *et al.* *Dev. Cell* **18**, 985–998 (2010).
8. Merino, M. M., Rhiner, C., Portela, M. & Moreno, E. *Curr. Biol.* **23**, 1300–1309 (2013).
9. Patel, P. H. & Edgar, B. A. *Semin. Cell Dev. Biol.* <http://dx.doi.org/10.1016/j.semcdb.2014.03.012> (2014).
10. Vincent, J.-P., Fletcher, A. G. & Baena-Lopez, L. A. *Nature Rev. Mol. Cell Biol.* **14**, 581–591 (2013).

This article was published online on 14 May 2014.

EARTH SCIENCE

Fertile fields for seismicity

An analysis of crustal uplift around California's San Joaquin Valley, caused by groundwater extraction, reveals that such removal leads to both seasonal and long-term unclamping of the nearby San Andreas Fault system. SEE LETTER P.483

PAUL LUNDGREN

Earthquakes are one of the most devastating manifestations of plate tectonics. Plate boundaries, such as the one between the Pacific and North American plates in California, are broken into multiple blocks separated by faults of varying sizes. Our understanding of earthquake size and occurrence is generally considered from the perspective of a fault's geological history^{1,2}, along with its present-day geodetic signature of surface deformation related to coupling of the fault at depth³. Earthquakes cause changes in stress along and across fault systems that either promote or deter seismicity⁴. Seismicity and fault stresses are considered largely on the basis of plate-boundary mechanics, without taking into account processes at Earth's surface, although the past decade has seen an increasing recognition that loading at the surface may be an important modulator of seismicity^{5–8}.

In this issue, Amos *et al.*⁹ (page 483) report an analysis of crustal uplift surrounding California's San Joaquin Valley that has resulted from water extraction. The study demonstrates that such anthropogenic effects can cause significant unclamping of the nearby San Andreas Fault system through flexure of the lithosphere (Earth's crust and uppermost mantle). Moreover, there is both a seasonal variation in and long-term promotion of seismicity associated with the water extraction. The latter may hasten the occurrence of future large earthquakes in the San Andreas Fault system (Fig. 1), although this study does not specify that one is imminent.

The San Joaquin Valley forms the southern half of California's large Central Valley, and lies between the Sierra Nevada mountains to the east and the Coast Ranges to the west. The valley is a significant source of a wide variety of crops, but with little precipitation it relies heavily on irrigation and thus has been subjected to long-term water withdrawal since

the late 1800s. This extraction of water exceeds its replenishment and results in considerable ground subsidence that is visibly evident and that has been measured geodetically within the valley itself. With the advent of the dense networks of continuous Global Positioning System (GPS) stations in the western United States, highly precise (less than 1 millimetre per year) vertical GPS velocities can now be obtained¹⁰.

In their study, Amos *et al.* analyse vertical GPS velocities across the San Joaquin Valley and find that a ring of crustal uplift surrounds the largely subsiding valley floor. By examining variations in the GPS time series, the authors isolate annual seasonal variations from each GPS site's multi-year trend. The seasonal variations relate to changes in precipitation mass in both the surrounding mountains and the San Joaquin Valley, whereas the multi-year trend reflects long-term water extraction in the valley. The result is a profile of crustal uplift from the Pacific coast across the Sierra Nevada that is consistent with a model for flexure of the lithosphere that is caused by changes in hydrologic loading across the valley. The authors fit this model to the observed vertical GPS velocities and estimate changes in the stresses (sliding and clamping) that drive faults adjacent to the valley, including the San Andreas Fault.

Amos *et al.* find that the flexure serves to unclamp and increase the sliding on the San Andreas Fault system. The unclamping of vertical faults such as the San Andreas occurs over the entire seismogenic portion of the fault (down to a depth of roughly 15 kilometres). Interestingly, the seasonal loading caused by water recharge in the San Joaquin Valley, and precipitation loading and unloading in the adjacent mountains, produces a Coulomb stress change (fault-driving stress) of around 1 kilopascal — a value that is of the same order of magnitude as the loading of faults due to nearby large earthquakes, and much greater than the long-term annual unloading signal from groundwater withdrawal. Over the long term (a 100-year timescale), the unloading generated by constant water extraction exceeds the seasonal effect and enhances the accumulation of stress between the occurrence of large earthquakes, bringing faults such as San Andreas closer to failure.

Amos and colleagues' analysis shows how anthropogenic effects, especially on a



Figure 1 | Aerial view of the San Andreas Fault along the Carrizo Plain, central California. Amos and colleagues' study⁹ suggests that seasonal and decadal water extraction in the nearby southern Central Valley stresses the San Andreas Fault system, causing variations in seismicity and promoting future earthquakes.

large spatial scale, can affect seismicity rates and potentially long-term earthquake hazard. Their analysis of vertical GPS velocities on the periphery of the San Joaquin Valley also shows that crustal uplift in the adjacent southern Sierra Nevada can be at least partly explained by groundwater withdrawal. Such uplift has previously been attributed to changes in glacial erosion during the Pleistocene geological epoch (between about 2,588,000 and 11,700 years ago), or to delamination (peeling away) of the Sierra Nevada lithosphere^{11,12}.

The ability to perform analyses such as that of Amos and colleagues relies on the existence of a dense, high-quality, continuous GPS network maintained over decades, coupled with an understanding of groundwater extraction. California's Central Valley is relatively large, with a simple geometry and faults that run parallel to the valley, thus allowing a quite simple interpretation of the crustal-uplift data and their effect on seismicity. The amount of scatter in the vertical GPS velocities requires that roughly 200 km of crustal-uplift

data be condensed into a single profile across the axis of the valley, ignoring second-order effects due to three-dimensional heterogeneity of the withdrawal pattern and fault-system geometry. The authors' flexure model fits the GPS velocities to within the data scatter, but does not use more realistic lithospheric properties, such as the viscous response of the lithosphere to temporally varying surface loading, that might allow better resolution of the fault-stress changes.

Furthermore, Coulomb stress change shows only those faults promoted to failure; it gives no information on the current state of stress on a fault or the physics of rupture initiation and dynamics¹³. Therefore, it is not a quantitative assessment of seismic hazard. As the number of dense, high-quality geodetic observations performed increase across the world, more analyses of this kind will be possible in regions of active tectonics and where there is large-scale anthropogenic interaction with the near-surface environment^{7,8}. This will be useful because the role of groundwater withdrawal in modulating seismicity and earthquake hazard is likely to become more significant as climate change perturbs precipitation patterns and as global population growth increases the demand for water. ■

Paul Lundgren is in the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.
e-mail: paul.r.lundgren@jpl.nasa.gov

1. Sieh, K. E. & Jahn, R. H. *Geol. Soc. Am. Bull.* **95**, 883–896 (1984).
2. van der Woerd, J. *et al. J. Geophys. Res.* **111**, B04407 (2006).
3. Savage, J. C. & Prescott, W. H. *J. Geophys. Res.* **83**, 3369–3376 (1978).
4. Stein, R. S. *Nature* **402**, 605–609 (1999).
5. Heki, K. *Earth Planet. Sci. Lett.* **207**, 159–164 (2003).
6. Luttrell, K., Sandwell, D., Smith-Konter, B., Bills, B. & Bock, Y. *J. Geophys. Res.* **112**, B08411 (2007).
7. Bettinelli, P. *et al. Earth Planet. Sci. Lett.* **266**, 332–344 (2008).
8. Fu, Y. & Freymueller, J. T. *J. Geophys. Res.* **117**, B03407 (2012).
9. Amos, C. B. *et al. Nature* **509**, 483–486 (2014).
10. Blewitt, G., Kreemer, C., Hammond, W. C. & Goldfarb, J. M. *J. Geodyn.* **72**, 11–24 (2013).
11. Small, E. E. & Anderson, R. S. *Science* **270**, 277–281 (1995).
12. Saleeby, J., Saleeby, Z. & Le Pourhiet, L. *Geosphere* **9**, 394–425 (2013).
13. Noda, H. & Lapusta, N. *Nature* **493**, 518–521 (2013).

This article was published online on 14 May 2014.

The role of senescent cells in ageing

Jan M. van Deursen

Cellular senescence has historically been viewed as an irreversible cell-cycle arrest mechanism that acts to protect against cancer, but recent discoveries have extended its known role to complex biological processes such as development, tissue repair, ageing and age-related disorders. New insights indicate that, unlike a static endpoint, senescence represents a series of progressive and phenotypically diverse cellular states acquired after the initial growth arrest. A deeper understanding of the molecular mechanisms underlying the multi-step progression of senescence and the development and function of acute versus chronic senescent cells may lead to new therapeutic strategies for age-related pathologies and extend healthy lifespan.

Cellular senescence is a process in which cells cease dividing and undergo distinctive phenotypic alterations, including profound chromatin and secretome changes, and tumour-suppressor activation^{1–6}. Hayflick and Moorhead first introduced the term senescence to describe the phenomenon of irreversible growth arrest of human diploid cell strains after extensive serial passaging in culture⁷. Later, this particular type of senescence (replicative senescence) was causally linked to telomere attrition, a process that leads to chromosomal instability and promotes tumorigenesis, supporting the original hypothesis that senescence guards against unrestricted growth of damaged cells^{7,8}. Subsequent studies have reinforced the importance of cellular senescence as a safeguard against cancer⁹. Emerging evidence indicates that the physiological relevance of cellular senescence extends beyond tumour suppression into biological processes such as embryonic development^{10–12}, wound healing¹³, tissue repair¹⁴ and organismal ageing^{15,16}. In fact, Hayflick and Moorhead initially postulated a role for replicative senescence in ageing, but until recently this theory remained untested⁷. The multifunctional nature of cellular senescence raises the question as to whether fundamentally different senescence mechanisms underlie these diverse biological roles. This Review focuses on this and other key emerging concepts in the senescence field, including ‘assisted’ cell cycling, multi-step senescence (or senescence progression), acute versus chronic senescence and senescence of post-mitotic cells. How these concepts relate to the role of senescent cells in ageing and age-related diseases and how the rapidly accruing new information could be exploited to clear detrimental senescent cell populations selectively to improve healthy lifespan are also discussed.

Causes and effector pathways of senescence

Research on the causes (or stresses), signalling networks and mechanisms underlying the various types of cellular senescence is still in its infancy and current insights are largely based on cell culture experiments. In addition to telomere erosion, several other tumour-associated stresses have been shown to induce a senescent growth arrest *in vitro*, including certain DNA lesions and reactive oxygen species (ROS)^{17–19}. What both these stresses have in common with telomere damage is that they activate the DNA damage response (DDR), a signalling pathway in which ATM or ATR kinases block cell-cycle progression through stabilization of p53 and transcriptional activation of the cyclin-dependent kinase (Cdk) inhibitor p21. Activated oncogenes are also prominent inducers of senescence. Oncogenic Ras acts through overexpression of Cdc6 and suppression of nucleotide metabolism, causing aberrant DNA replication, formation of double stranded DNA breaks (DSBs) and activation of the DDR pathway^{20,21}. However, senescence caused by E2F3 activation or c-Myc inhibition is

DDR-independent and involves p19^{Arf} and p16^{Ink4a} (refs 17, 22). BRAF (V600E) is also DDR-independent and induces senescence through a metabolic mechanism involving upregulation of mitochondrial pyruvate dehydrogenase (PDH; Fig. 1)²³. Several other studies underscored that senescence is closely linked to profound metabolic changes^{24,25}. Furthermore, various tumour suppressors trigger a senescent growth arrest when inactivated, including RB, PTEN, NF1 and VHL^{17,26}. Of these, RB inactivation engages the DDR²⁶, whereas the others are DDR-independent and act through p19^{Arf} and p16^{Ink4a}. A notable species-specific difference is that senescence pathways of murine cells are more dependent on p19^{Arf} than senescence in human cells²⁷.

Prolonged exposure to interferon- β also induces senescence, demonstrating that chronic mitogenic signalling outside the context of neoplastic transformation can stimulate senescence²⁸. Other, less broadly studied inducers of senescence include epigenetic, nucleolar and mitotic spindle stresses (Fig. 1). For example, genome-wide chromatin decompression by exposure to histone deacetylase inhibitors triggers senescence via a p21-dependent mechanism²⁹. A key target of epigenetic stressors that promote senescence may be the *INK4a/ARF* locus, which in proliferating cells is repressed by polycomb group-mediated H3K27 methylation and H2A-K119 ubiquitination³⁰. Nucleolar stress caused by RNA polymerase I inhibitors triggers a robust p53-mediated senescence response³¹. Senescence can also be elicited by suboptimal expression of proteins implicated in spindle formation or mitotic checkpoint control, including human TACC3 and murine BubR1, Bub3 and Rae1, all of which engage p53 and p21 independently of the DDR, often in combination with p16^{Ink4a} (refs 15, 32, 33). It is highly likely that additional stressors and mechanisms that drive cells into senescence will be uncovered given the rapidly evolving nature of the field. Production of proinflammatory cytokines and chemokines is emerging as a common feature of senescent cells irrespective of the senescence-inducing stressor or mechanism (Fig. 1).

Although the relative contributions of the p53–p21 and p16^{Ink4a}–RB effector pathways to the initial growth arrest can vary depending on the type of stress, both may ultimately become engaged upon sustained senescence. For example, DNA damage initially halts cell-cycle progression through p53-mediated induction of p21, but if lesions persist, this activates p16^{Ink4a} through p38-MAPK-mediated mitochondrial dysfunction and ROS production^{34,35}. The extent to which effector mechanisms of *in vitro* senescence apply to *in vivo* senescence has not been tested extensively. Fat, skeletal muscle and eye of BubR1 progeroid mice have elevated levels of p19^{Arf}, p53, p21 and p16^{Ink4a} and are subject to precocious functional decline^{15,36}. Genetic experiments using knockout strains for each of these tumour suppressors that dissected how senescent cells accumulate in

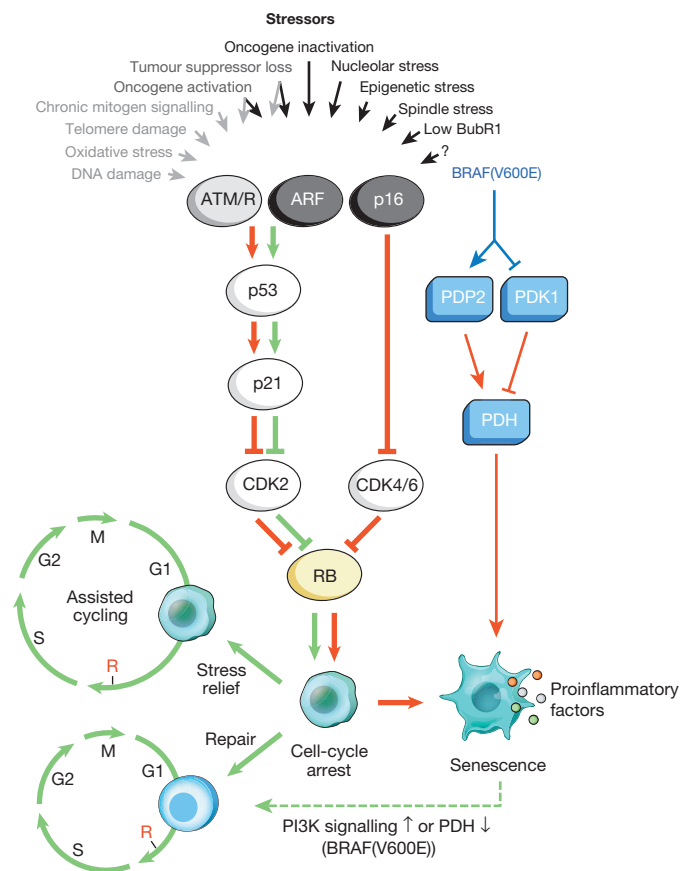


Figure 1 | Senescence-inducing stimuli and main effector pathways.

A variety of cell-intrinsic and -extrinsic stresses can activate the cellular senescence program. These stressors engage various cellular signalling cascades but ultimately activate p53, p16^{Ink4a}, or both. Stress types that activate p53 through DDR signalling are indicated with grey text and arrows (ROS elicit the DDR by perturbing gene transcription and DNA replication, as well as by shortening telomeres). Activated p53 induces p21, which induces a temporal cell-cycle arrest by inhibiting cyclin E–Cdk2. p16^{Ink4a} also inhibits cell-cycle progression but does so by targeting cyclin D–Cdk4 and cyclin D–Cdk6 complexes. Both p21 and p16^{Ink4a} act by preventing the inactivation of Rb, thus resulting in continued repression of E2F target genes required for S-phase onset. Upon severe stress (red arrows), temporally arrested cells transition into a senescent growth arrest through a mechanism that is currently incompletely understood. Cells exposed to mild damage that can be successfully repaired may resume normal cell-cycle progression. On the other hand, cells exposed to moderate stress that is chronic in nature or that leaves permanent damage may resume proliferation through reliance on stress support pathways (green arrows). This phenomenon (termed assisted cycling) is enabled by p53-mediated activation of p21. Thus, the p53–p21 pathway can either antagonize or synergize with p16^{Ink4a} in senescence depending on the type and level of stress. BRAF(V600E) is unusual in that it establishes senescence through a metabolic effector pathway. BRAF(V600E) activates PDH by inducing PDP2 and inhibiting PDK1 expression, promoting a shift from glycolysis to oxidative phosphorylation that creates senescence-inducing redox stress. Cells undergoing senescence induce an inflammatory transcriptome regardless of the senescence inducing stress (coloured dots represent various SASP factors). Red and green connectors indicate ‘senescence-promoting’ and ‘senescence-preventing’ activities, respectively, and their thickness represents their relative importance. The dashed green connector denotes a ‘senescence-reversing’ mechanism.

these tissues and contribute to their deterioration, established that p16^{Ink4a} is an effector of senescence and ageing¹⁵. However, in contrast to *in vitro* findings, p19^{Arf}, p53 and p21 prevented senescence and age-related pathologies *in vivo*^{15,36}. These unexpected findings led to the concept of ‘assisted’ cell cycling (analogous to assisted living) in which ageing cells, coping with an increasing burden of macromolecular damage and other chronic stresses,

manage to retain their proliferative potential for a while by extending cell-cycle duration in a p21-dependent manner to provide extra time to mend cellular disabilities through engagement of compensatory mechanisms or repair (Fig. 1). Although this concept is supported by the observation that the cell-cycle time of cultured primary human cells markedly increases with passaging³⁷, it clearly requires more validation and generalization. Inactivation of p21 improves stem cell function in intestinal crypts and bone marrow in mutant mice with short telomeres³⁸, indicating that in situations where irreparable damage produces a sustained and robust p53 response, p21 acts to promote tissue deterioration by executing senescence.

In vitro studies of cellular senescence have traditionally been performed using a single senescence-inducing stimulus (that is, high-dose radiation or oncogenes; Fig. 1). However, in the context of organismal ageing, individual cells experience multiple cellular pressures, including various kinds of genotoxic, proteotoxic and mitotic stresses^{3,39}. Thus, to advance our understanding of these processes, it will be imperative to examine how combinations of diverse senescence-promoting stressors impact the actions of the various downstream effector pathways and the characteristics of the resulting senescent phenotypes. Furthermore, while cellular senescence is well recognized as an *in vivo* tumour suppressive mechanism, its irreversibility remains a topic of debate. However, compelling new evidence indicates that BRAF(V600E) oncogene-induced senescence (OIS) can be reversed by activation of phosphatidylinositol 3-kinase (PI3K) or inhibition of PDH (Fig. 1)^{23,40}. In addition, senescent cells have been successfully dedifferentiated into pluripotent stem cells⁴¹.

Senescence is a multi-step evolving process

Until recently, senescence was viewed as a static endpoint. However, several recent observations support the hypothesis that senescence can be a highly dynamic, multi-step process, during which the properties of senescent cells continuously evolve and diversify, much like tumorigenesis but without cell proliferation as a driver (Fig. 2)^{42–44}. The initiating step is the transition of temporal to stable cell-cycle arrest, which typically involves prolonged inhibition of Cdk–cyclin activity by p21, p16^{Ink4a}, or both. A change in p53 expression from intermittent to continuous may be a critical event in the transition from temporal to persistent growth arrest⁴⁵.

For the progression to full senescence, it seems that lamin B1 downregulation triggers both global and local modifications in chromatin methylation^{46–48} (Fig. 2). Some mammalian cell types form regions of highly condensed chromatin called senescence-associated heterochromatin foci (SAHFs)^{49–51}. SAHFs, which are enriched in chromatin modifications such as S83-HP1γ, HIRA, ASF1, macroH2A, H3K9me3 and γH2AX, sequester genes implicated in cell-cycle control, a phenomenon that seems to reinforce the senescence-associated growth arrest. Decondensation of (peri)centromeric satellite heterochromatin has been identified as a universal hallmark of senescence that precedes SAHF formation⁵². Senescence-related chromatin remodeling leads to profound transcriptional changes^{48,53,54}. Among the assortment of upregulated genes is a prominent subset of genes that encode secreted proteins, including cytokines and chemokines with proinflammatory properties, as well as various growth factors and proteases that together alter tissue structure and function. Collectively, these factors are referred to as the senescence-associated secretory phenotype (SASP)^{55,56} or senescence-messaging secretome (SMS)⁵⁷. The SASP is one of the key characteristics that distinguish senescent cells from quiescent, terminally differentiated, and other types of non-proliferating cells (Fig. 2). In certain cases, the SASP is dependent on persistent DNA damage signalling⁵⁶, such as that created through a positive feedback loop between DDR signalling and ROS⁵⁵. This loop was uncovered using human fibroblast lines in which dominant-negative TRF2 or high dose ionizing radiation induced telomere-dependent and telomere-independent DDR signalling, respectively. Both types of DDR signalling were found to cause mitochondrial dysfunction and production of ROS which led to new DNA damage and continued DDR signalling. Consistent with this, primary human fibroblasts overexpressing p16^{Ink4a} or p21 undergo senescent growth arrest but fail to activate the DDR and do not produce a SASP⁵⁸. Importantly, several senescence-inducing stimuli

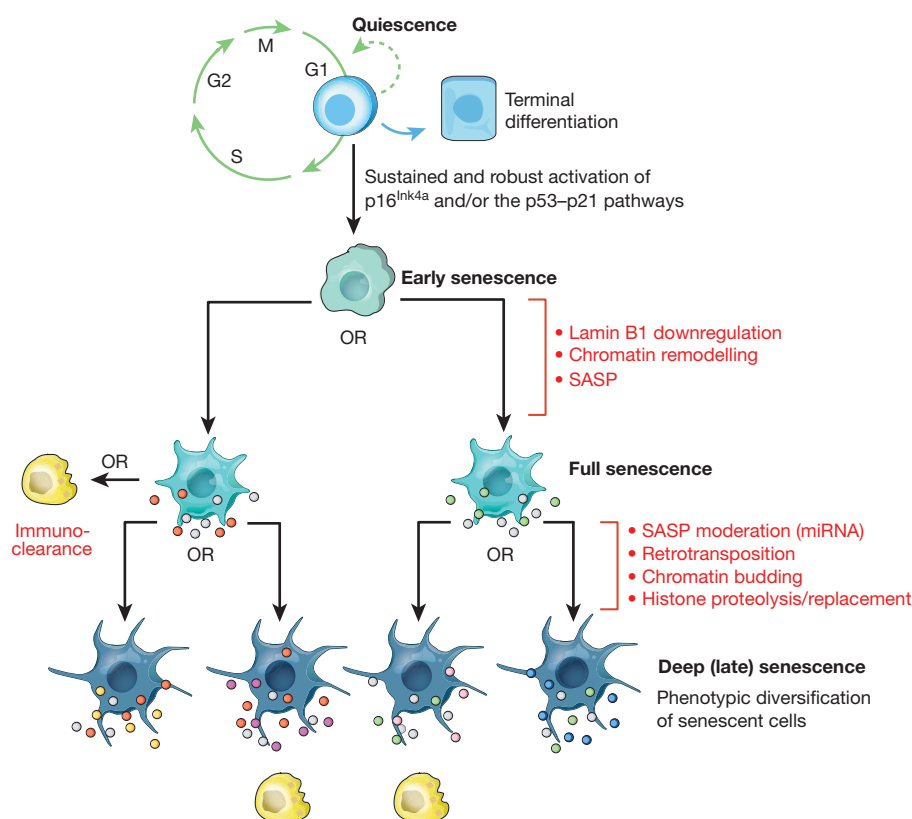


Figure 2 | Hypothetical multi-step senescence model. Mounting evidence suggests that cellular senescence is a dynamic process driven by epigenetic and genetic changes. The initial step represents the progression from a transient to a stable cell-cycle arrest through sustained activation of the p16^{Ink4a} and/or p53–p21 pathways. The resulting early senescent cells progress to full senescence by downregulating lamin B1, thereby triggering extensive chromatin remodelling underlying the production of a SASP. Certain components of the SASP are highly conserved (grey dots), whereas others may vary depending on cell type, nature of the senescence-inducing stressor, or cell-to-cell variability in chromatin remodelling (red and green dots). Progression to deep or late senescence may be driven by additional genetic and epigenetic changes, including chromatin budding, histone proteolysis and retrotransposition, driving further transcriptional change and SASP heterogeneity (yellow, magenta, pink and blue dots). It should be emphasized that although the exact nature, number and order of the genetic and epigenetic steps occurring during senescent cell evolution are unclear, it is reasonable to assume that the entire process is prone to SASP heterogeneity. The efficiency with which immune cells (yellow) dispose of senescent cells may be dependent on the composition of the SASP. Interestingly, the proinflammatory signature of the SASP can fade due to expression of particular microRNAs late into the senescence program, thereby perhaps allowing evasion of immuno-clearance⁹⁹.

produce SASPs independent of DNA damage (Fig. 1), implying the existence of DDR-independent mechanisms^{11,12,23}.

Given the intricate nature of the SASP, it is not surprising that senescent cells impact various biological processes that involve paracrine signalling, including cell proliferation, angiogenesis⁵⁹, inflammation⁶⁰, epithelial-to-mesenchymal transition (EMT)⁶¹, wound healing¹³, and other types of tissue repair¹⁴. Several SASP components, including IL-6, IL-8, WNT16B and GRO α , also act in an autocrine fashion in the context of OIS, presumably to help establish a persistent growth arrest^{62–64}. Importantly, SASP factors vary in distinct cell types and under different senescence-inducing stressors⁵⁵. This plasticity within SASP composition predicts variability with respect to the biological processes impacted by different kinds of senescent cells (Fig. 2). Proinflammatory cytokines and chemokines are among the SASP components that are highly conserved across cell types and senescence-inducing stimuli⁵⁵, suggesting that attracting immune cells and inducing local inflammation are common properties of senescent cells. However, accumulation of senescent cells is not always accompanied by immune cell infiltration and inflammation, as exemplified by melanocytic nevi⁶⁵.

Cultured cells usually reach senescence within several weeks after exposure to senescence-inducing stressors, but remain viable for months thereafter⁴². Senescent cells continue to evolve even after extended periods of culture, thereby progressing to a stage that has been termed ‘deep’ or ‘late’ senescence (Fig. 2). This phenomenon is evidenced by a dramatic increase in the transcription of transposable elements, including members of the L1, ALU and SVA transposon families, which occur several months after senescence onset^{42,43}. These newly synthesized retrotransposon transcripts can indeed engage in active transposition and accumulate in late-senescent cell genomes. Increased retrotransposon activity is associated with senescence-associated opening of gene-poor heterochromatic regions where these elements reside⁴². A second process driving continued change in senescent cells is characterized by the extrusion of chromatin into the cytoplasm, resulting in the formation of cytoplasmic chromatin fragments (CCFs)⁴⁴. CCFs are strongly positive for H3K27me3 and γ H2AX, contain DNA, and are processed via lysosome-mediated proteolysis, resulting in overall histone

loss. Both retrotransposon activation and chromatin budding are examples of continued genomic and epigenomic remodelling in senescent cells. Both processes seem random in nature and are therefore likely to drive transcriptome diversity among senescent cells, even in those resulting from a common stressor. Because transcriptional activity is a key determinant of secretome composition, it is reasonable to assume that the SASPs of these cells diversify to some degree through these changes. With the concept of senescence progression in cultured cells solidifying, it becomes essential to validate its *in vivo* importance.

Acute versus chronic senescence

The diverse nature of the processes in which senescence has been implicated, ranging from embryonic development¹⁰ to wound healing¹³, tissue repair¹⁴, cancer and ageing¹⁶, raises the question of whether the properties of the senescent cells involved in these activities are fundamentally different. If so, what would be the underlying molecular mechanisms? Valuable clues to these questions can be inferred from the apparent differences in senescence kinetics between these processes. In the context of ageing, cells chronically accumulate macromolecular damage and may become increasingly dependent on cell-cycle checkpoints and stress-relief mechanisms to retain proliferative potential (assisted cycling)^{15,66–68}. Ultimately, more and more of these cells may stably arrest and transition into a senescent state, referred to here as chronic senescence (Fig. 3).

Chronic senescence is different from a process like wound healing, by which upon wound closure, myofibroblasts suddenly undergo senescence (acute senescence) to limit excessive fibrosis at the site of injury¹³ (Fig. 3). Myofibroblast senescence here is induced by the extracellular matrix protein CCN1, which acts through integrin $\alpha 6 \beta 1$ and HSPS-mediated activation of the RAC1-dependent NADPH oxidase 1 to produce a robust and sustained accumulation of ROS¹³. Senescent myofibroblasts limit fibrosis by promoting the degradation of matrix components through the SASP factors they produce. Importantly, acute myofibroblast senescence also has a role in repairing damaged organs such as the liver and thus is likely to represent a more common mechanism to limit fibrosis^{14,69}.

Similar to skin repair and recovery from liver injury, senescent cell induction is also acute and spatiotemporally controlled in uterine neo-vascularization, a developmental process that takes place at the site of embryo implantation to supply the embryo with maternal blood¹⁰. HLA-G secreted by embryonic trophoblast cells induces senescence in nearby natural killer cells, which then start to produce SASP components to promote local angiogenesis and vascular remodelling. Senescence has also been identified as a prominent mechanism for remodelling of various tissues during mouse embryogenesis, including the mesonephros, the endolymphatic sac, the apical ectodermal ridge, and the neural roof plate^{11,12}. Developmental senescence is p21-dependent, but p53- and DDR-independent, and shares several features with OIS, including a common gene expression signature and senescence-associated β -galactosidase activity. Interestingly, OIS itself is triggered by a single defined stimulus and established with fast kinetics, which would qualify it as acute senescence. However, cells undergoing OIS are not always cleared by the immune system. For instance senescent cells in human melanocytic nevi are highly persistent.

Thus, senescence induction in tissue repair and development seems to be a scheduled or programmed process triggered by specific stimuli that target particular types of cells (Fig. 3). In contrast, during ageing-related senescence, the switch from temporal to persistent cell-cycle arrest appears unscheduled and stochastic in nature, probably involving the combined effects of distinct senescence-inducing stressors acting simultaneously on a cell. The kinetics and efficiency of senescent-cell clearance may constitute another key difference between acute and chronic senescence. During repair and embryogenesis, disposal of senescent cells seems very efficient and under strict temporal control^{11–14}. Conversely, ageing-related senescent cells, may be more persistent due to deterioration of the immune system with ageing^{70,71}, but further experiments are needed to refine our understanding of the relationship between senescence and the (ageing) immune system.

Senescence induced by chemotherapeutics or radiation in the context of cancer treatment may be a combination of acute and chronic senescence (Fig. 3). Acute senescence would apply to cells that generate a sustained

DDR as a rapid response to overwhelming genomic damage⁷². In contrast, chronic senescence would pertain to cells experiencing mild genotoxic stress that can initially be managed through engagement of stress support pathways. However, as further macromolecular damage occurs over time, these cells may eventually transition from a pre-senescent state to a persistent cell cycle arrest⁷³. An interesting untested hypothesis is that cancer therapy-induced senescence contributes to accelerated tissue and organ deterioration in cancer survivors⁷⁴.

Senescence of post-mitotic cells

Most cells in mammals are post-mitotic and the question that has been raised is whether these cells can obtain key characteristics of senescent cells. Post-mitotic neurons in various parts of human and mouse brains are known to accumulate high amounts of DNA damage¹⁸. Recent research has revealed that these neurons exhibit several additional senescence-associated properties, including heterochromatinization, synthesis of proinflammatory interleukins, and high senescence-associated β -galactosidase activity⁷⁵. As with mitotic cells that undergo senescence in response to sustained DNA damage, these phenotypes develop in a p21-dependent manner, further associating them with senescence (Fig. 4). Senescence-like features have also been reported for adipocytes of mice on a high-fat diet⁷⁶, suggesting that post-mitotic cell senescence may be a broader phenomenon. It will be important to confirm that these terminally differentiated cells produce a SASP that negatively impacts the functionality of neighbouring cells (Fig. 4), and to explore whether differentiated cells with senescent cell properties accumulate in tissues other than brain and fat.

Senescence in ageing and age-related disease

The absence of senescence specific markers has hampered efforts to characterize senescent cells that accumulate *in vivo* in tissues and organs. The most reliable *in situ* detection methods that are currently available screen for multiple semi-selective senescent cell characteristics. These include, but are not limited to, high levels of p16^{Ink4a}, p21, macroH2A, IL-6, phosphorylated

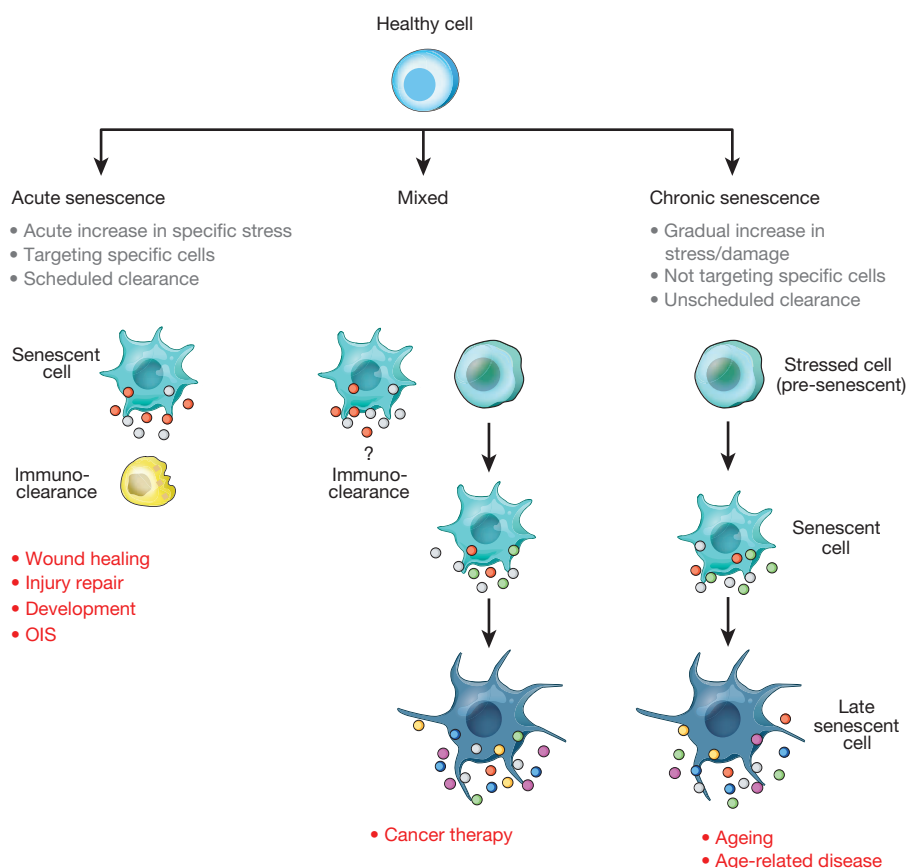


Figure 3 | Acute and chronic senescent cells. The conceptual model in which senescent cells are subdivided into two main classes based on kinetics of senescence induction and functionality. Acute senescent cells seem to mostly be part of tightly orchestrated biological processes (that is, wound healing, tissue repair, embryonic development) to halt expansion of certain cells and/or produce a SASP with defined paracrine functions. Acute senescence is induced through cell-extrinsic stimuli that target a specific population of cells in the tissue. Acute senescent cells self-organize their elimination through SASP components that attract various types of immune cells. Induction of chronic senescence occurs after periods of progressive cellular stress or macromolecular damage when tarry cycling transitions into a stable cell-cycle arrest. Chronic senescence is not programmed and does not seem to target specific cell types. Conceivably, owing to age-related immunodeficiency or production of less proinflammatory SASPs, immune cells may inefficiently eliminate chronic senescent cells, allowing continuation of multi-step senescence. Senescence induced during cancer therapy may initially be acute and later chronic in nature.

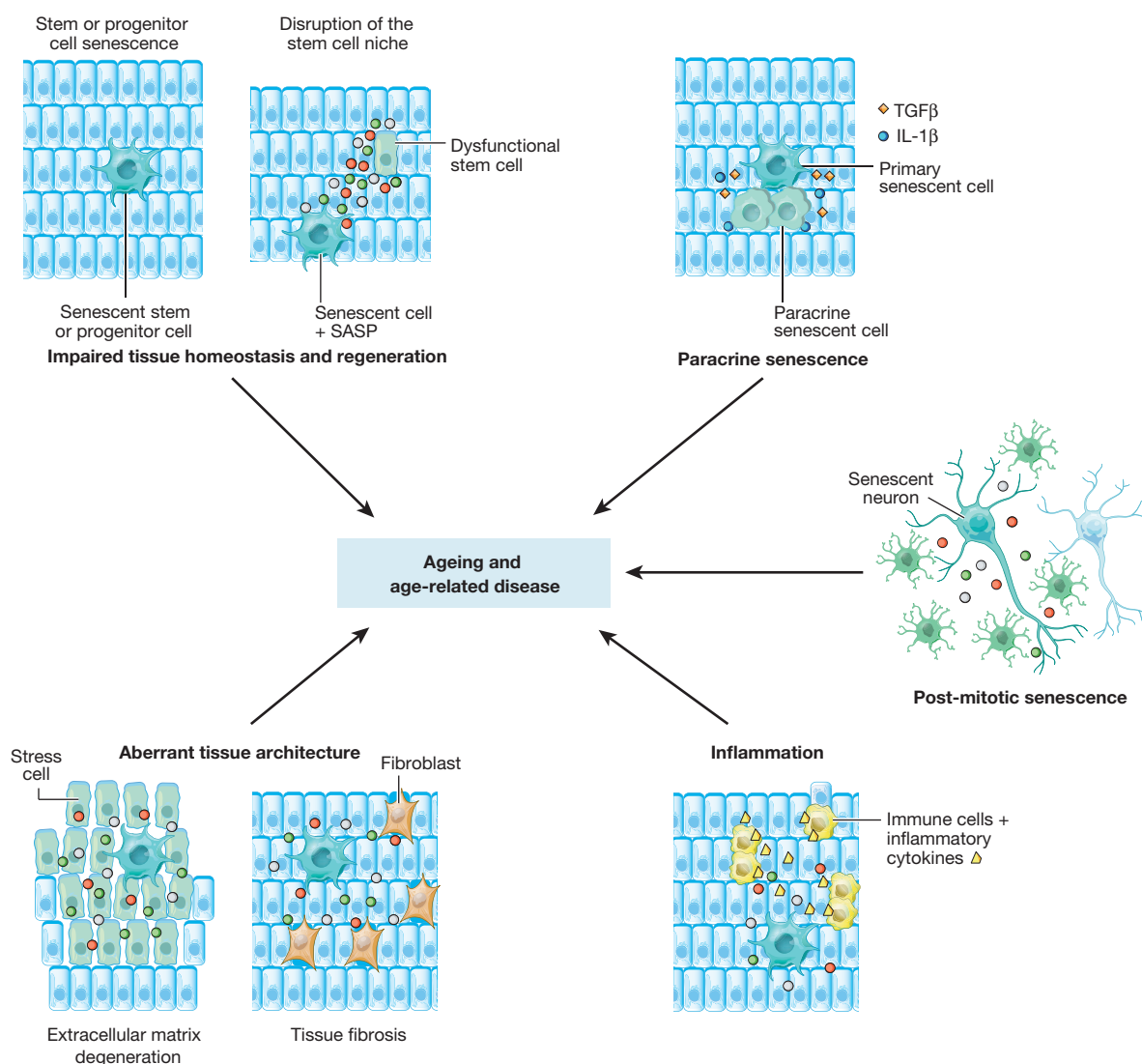


Figure 4 | Mechanisms of tissue and organ deterioration by cellular senescence. Cellular senescence is thought to contribute to age-related tissue and organ dysfunction and various chronic age-related diseases through various mechanisms. In a cell-autonomous manner, senescence acts to deplete the various pools of cycling cells in an organism, including stem and progenitor cells. In this way, senescence interferes with tissue homeostasis and regeneration, and lays the groundwork for its cell-non-autonomous detrimental actions involving the SASP. There are at least five distinct paracrine mechanisms by which senescent cells could promote tissue dysfunction,

including perturbation of the stem cell niche (causing stem cell dysfunction), disruption of extracellular matrix, induction of aberrant cell differentiation (both creating abnormal tissue architecture), stimulation of sterile tissue inflammation, and induction of senescence in neighbouring cells (paracrine senescence). An emerging yet untested concept is that post-mitotic, terminally differentiated cells that develop key properties of senescent cells might contribute to ageing and age-related disease through the same set of paracrine mechanisms.

p38 MAPK, DSBs, and senescence-associated β -galactosidase activity. The use of such methods has provided convincing evidence that senescent cells indeed accumulate in tissues of humans, primates, and rodents with age^{77–81}, as well as at sites of tissue injury and remodelling^{10–14}. Furthermore, cells with senescent cell properties can be found in the affected tissues of patients with age-related diseases such as osteoarthritis, pulmonary fibrosis, atherosclerosis, and Alzheimer's disease^{1,82}.

Studies on the relevance of *in vivo* senescence in ageing and age-associated diseases have also been complicated by the fact that key downstream effectors of senescence-inducing stressors such as p16^{Ink4a} and p53 are tumour suppressors that, when disrupted in mice, cause death from cancer at a young age⁸³. p53 provides additional complications to these analyses as, unlike p16^{Ink4a}, this transcription factor also mediates apoptosis, making it difficult to assign potential ageing-related phenotypic changes in p53-null animals to the senescence program^{84–86}. Two consecutive studies in BubR1 progeroid mice, in which p16^{Ink4a}-positive senescent cells were targeted in different ways provided the first direct support for Hayflick and Moorhead's

early concept that senescent cells drive age-related pathologies^{1,7,15,16}. In the first study, genetic inactivation of p16^{Ink4a} prevented the formation of senescent cells in skeletal muscle, eye and fat, significantly attenuating the onset of age-related pathologies in these tissues¹⁵. In the subsequent study, which produced a phenocopy of genetic p16^{Ink4a} ablation, p16^{Ink4a}-positive senescent cells were allowed to accumulate but were consistently eliminated from weaning age onwards by the use of a transgene, termed INK-ATTAC, that selectively induced apoptosis in these cells upon administration of the synthetic drug AP20187 (ref. 16). Late-life clearance of senescent cells attenuated progression of already-established age-related disorders in skeletal muscle and fat, yet was unable to revert them.

An important question that needs to be addressed is how senescence promotes age-related tissue dysfunction (Fig. 4). One scenario is that senescence contributes to the overall decline in tissue regenerative potential that occurs with ageing. This idea is supported by the observation that progenitor cell populations in both skeletal muscle and fat tissue of BubR1 progeroid mice are highly prone to cellular senescence³⁶. In addition to

acting on stem cells in a cell-autonomous fashion by establishing a persistent growth arrest, senescence could act to disrupt the local stem-cell niche non-autonomously through the SASP^{87–89} (Fig. 4). Although this concept remains to be tested *in vivo*, the profound negative impact that the aged cellular microenvironment has on stem cell functionality is underscored by the discovery that the regenerative potential of old stem cells markedly improves when exposed to a young systemic environment via parabiotic pairing^{87,90}.

Other SASP-based mechanisms may also contribute to tissue dysfunction. For example, proteases chronically secreted by senescent cells may perturb tissue structure and organization by cleaving membrane-bound receptors, signalling ligands, extracellular matrix proteins or other components in the tissue microenvironment^{55,91} (Fig. 4). In addition, other SASP components, including IL-6 and IL-8, may stimulate tissue fibrosis in certain epithelial tissues by inducing EMT^{61,91} (Fig. 4). Chronic tissue inflammation, which is characterized by infiltration of macrophages and lymphocytes, fibrosis and cell death, is associated with ageing and has a causal role in the development of various age-related diseases⁶⁰. One idea, which remains untested, is that senescent cells that accumulate with ageing and that are present at sites of age-related pathologies promote this type of inflammation through the proinflammatory growth factors, cytokines and chemokines they secrete (Fig. 4). These may include GM-CSF, GRO α , IL-1, IL-6, IL-8, macrophage inflammatory proteins (MIPs), as well as monocyte chemo-attractant proteins (MCPs)^{1,49,55}. Together with matrix metalloproteinases, proinflammatory SASP components are thought to create a tissue microenvironment that promotes survival, proliferation and dissemination of neoplastic cells, which may explain, at least in part, why cancer rates markedly increase beyond middle age^{1,80,84}. Finally, the SASP may intensify age-related tissue deterioration through paracrine senescence, a recently discovered mechanism by which senescent cells spread the senescence phenotype to healthy neighbouring cells through secretion of IL-1 β , TGF β and certain chemokine ligands (Fig. 4)^{92,93}.

Why senescent cells accumulate in tissues and organs with age is another key open question. One possibility is that the rate with which senescent cells are produced might increase over time. In support of this idea, several investigations have demonstrated that various stimuli that induce senescence increase with ageing^{32,94,95}. If combined cellular stresses were to drive senescence, it would take a long time for these to accumulate. Alternatively, the efficiency with which senescent cells get eliminated may decrease with ageing. In fact, senescent cells can be killed and disposed of by immune cells, as was elegantly demonstrated in mice that undergo senescence in the context of liver fibrosis and hepatocellular carcinogenesis^{14,96}. These observations raised the possibility that senescent cells are armed with a self-elimination program that proceeds by attracting both adaptive and innate immune cells, including T cells, macrophages and natural killer cells, through the secretion of proinflammatory cytokines and chemokines^{60,96,97}. This program is likely to be affected in ageing humans and rodents, in which the immune system undergoes a complex series of changes in both the innate and adaptive immunity that culminate in age-associated immunodeficiency⁷¹. This may include a reduced efficiency of senescent-cell clearance. Indeed, age-related haematopoietic stem cell dysfunction compromises the immune system and may thus be an important contributor to the late-life systemic increases in senescent cells^{70,98}. Furthermore, given that senescence may be a dynamic process rather than a static state, it is conceivable that the ability to self-eliminate through immune cells becomes compromised as senescent cells evolve.

Another important consideration is whether chronic and acute senescence can both play a role in ageing and age-related disease. Acute senescent cells are part of tightly orchestrated biological processes in which they have narrowly defined roles and a temporal presence both dictated by the composition of the SASP, limiting their ability to accumulate with ageing. In contrast, chronic senescent cells which develop after a prolonged period of gradually increasing cellular stresses are expected to exhibit high SASP heterogeneity due to a more complex and diverse spectrum of effector pathways involved in establishing this type of longer lasting senescent state. SASP heterogeneity may therefore be a mechanism to create subsets of

senescent cells that are highly resistant to immune clearance and drive tissue degeneration.

Senescent-cell clearance and future directions

Studies in BubR1 progeroid mice provided proof-of-principle that clearance of senescent cells can delay age-related degenerative pathologies¹⁶. This, together with the lack of overt detrimental side effects associated with long-term clearance in BubR1 mutant mice, suggests that targeting senescent cells for destruction might be an effective therapeutic strategy for treatment of age-related diseases or improvement of healthy lifespan. Evidently, sophisticated approaches developed for selective eradication of cancer cells provide an invaluable blueprint for the development of molecular-targeted therapies against senescent cells. The problem of drug resistance that has plagued the cancer field is unlikely to apply to senolytic agents because rare residual senescent cells that remain after treatment would not be able to amplify through division.

However, although senescent-cell removal represents an attractive therapeutic avenue, there are many unknowns and potential pitfalls along this route. For example, our current knowledge about the rates and spatio-temporal patterns that drive the accumulation of senescent cells in both humans and animal models during normal ageing and in age-related diseases is limited. Another gap in knowledge relates to the degree of phenotypic heterogeneity (that is, SASP composition) between senescent cells that accumulate *in vivo*, not only between the acute and chronic senescent cells but also within these two classes. Also, it will be imperative to determine the impact of senescent cell clearance on the health and lifespan of normal mice, particularly now that evidence is mounting that senescence is beneficial for tissue development and repair.

Another important consideration is whether the mouse is a reliable model for recapitulating the physiological effects of senescence cell accumulation and clearance that occurs in humans. One prominent senescence-inducing stressor, telomere attrition, is specific to humans, and may be responsible for a higher baseline level of senescence in our species. If similar to telomere attrition, other senescence-inducing mechanisms were indeed more prevalent in humans than in mice, the therapeutic effect of senescent cell clearance might be even more robust than in mice. Conversely, we should consider that, in spite of its potential beneficial effects, the removal of high percentages of senescent cells could have undesirable outcomes to human health by triggering atrophy and tissue dysfunction. In addition, the most illuminating experiments in mice relied on targeting p16^{Ink4a}-positive cells for elimination, a population that probably represents only a subset of cells undergoing senescence. Thus, the effects of clearing either p16^{Ink4a}-negative senescent cells or, collectively, all senescent cells remain to be determined. In addition, cultured human and mouse cells differentially rely on the p53 and p16^{Ink4a} pathways for induction of senescence²⁷, underscoring the need for carefully validating information from mouse studies before extrapolating its results to the human situation. We should also consider that although the existence of multiple senescent-cell subtypes offers at the same time challenges and opportunities, targeting each of these populations separately might have either beneficial or detrimental effects.

The field of experimental therapeutics as it relates to senescence is a nascent yet promising area of investigation. Besides small molecules that target senescent cells, a potentially promising and straightforward bio-therapeutic approach would be to activate or reinforce the immune response against senescent cells. This approach will require a deeper understanding of the extent to which the immune system disposes of senescent cells as well as the molecular and cellular mechanisms underlying this process. Potential caveats are that immunodeficiency is not a main contributor to age-related increases in senescent cells, or that chronically senescent cells are enriched in SASP components that provide resistance to immune clearance⁹⁹. Approaches exploiting the adaptive immune system to mount effective immunity against senescent cells may have limited feasibility owing to a potential lack of senescent-cell-specific antigens.

Undoubtedly, the next decade will see a tremendous expansion of data on the mechanisms, characteristics and functions of *in vivo* senescence, as

well as the use of this information to ameliorate human age-related diseases and promote healthy lifespan.

Received 6 October 2013; accepted 3 March 2014.

1. Campisi, J. Aging, cellular senescence, and cancer. *Annu. Rev. Physiol.* **75**, 685–705 (2013).
2. Kuilman, T., Michaloglou, C., Mooi, W. J. & Peeper, D. S. The essence of senescence. *Genes Dev.* **24**, 2463–2479 (2010).
3. López-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
4. Adams, P. D. Healing and hurting: molecular mechanisms, functions, and pathologies of cellular senescence. *Mol. Cell* **36**, 2–14 (2009).
5. Newgard, C. B. & Sharpless, N. E. Coming of age: molecular drivers of aging and therapeutic opportunities. *J. Clin. Invest.* **123**, 946–950 (2013).
6. Tchikonia, T., Zhu, Y., van Deursen, J., Campisi, J. & Kirkland, J. L. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities. *J. Clin. Invest.* **123**, 966–972 (2013).
7. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).

A pioneering study that introduced the term senescence to describe the phenomenon of permanent growth arrest of primary human cells after extensive serial passaging in culture.

8. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
9. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).

The study that established the concept of oncogene-induced senescence.

10. Rajagopalan, S. & Long, E. O. Cellular senescence induced by CD158d reprograms natural killer cells to promote vascular remodeling. *Proc. Natl Acad. Sci. USA* **109**, 20596–20601 (2012).
11. Muñoz-Espin, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
12. Storer, M. *et al.* Senescence is a developmental mechanism that contributes to embryonic growth and patterning. *Cell* **155**, 1119–1130 (2013).

References 10, 11, 12 are studies demonstrating that senescence has a central role in tissue remodelling during embryogenesis.

13. Jun, J. I. & Lau, L. F. The matricellular protein CCN1 induces fibroblast senescence and restricts fibrosis in cutaneous wound healing. *Nature Cell Biol.* **12**, 676–685 (2010).
14. Krizhanovsky, V. *et al.* Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008).

References 14, 15 show that senescent cells that are produced after tissue damage act to curb fibrosis.

15. Baker, D. J. *et al.* Opposing roles for p16^{INK4a} and p19^{Arf} in senescence and ageing caused by BubR1 insufficiency. *Nature Cell Biol.* **10**, 825–836 (2008); corrigendum **14**, 649 (2012).

A study that demonstrated a causal link between cellular senescence and age-related tissue deterioration, and the concept of assisted cycling.

16. Baker, D. J. *et al.* Clearance of p16^{INK4a}-positive senescent cells delays ageing-associated disorders. *Nature* **479**, 232–236 (2011).

A study showing that clearance of p16^{INK4a}-positive senescent cells can delay age-related degenerative pathologies in a progeroid mouse model.

17. Nardella, C., Clohessy, J. G., Alimonti, A. & Pandolfi, P. P. Pro-senescence therapy for cancer treatment. *Nature Rev. Cancer* **11**, 503–511 (2011).
18. Sedelnikova, O. A. *et al.* Senescent human cells and ageing mice accumulate DNA lesions with unreparable double-strand breaks. *Nature Cell Biol.* **6**, 168–170 (2004).
19. von Zglinicki, T. Oxidative stress shortens telomeres. *Trends Biochem. Sci.* **27**, 339–344 (2002).
20. Aird, K. M. *et al.* Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence. *Cell Rep.* **3**, 1252–1265 (2013).
21. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
22. Lazzarini, D., Denchi, E., Attwooll, C., Pasini, D. & Helin, K. Deregulated E2F activity induces hyperplasia and senescence-like features in the mouse pituitary gland. *Mol. Cell Biol.* **25**, 2660–2672 (2005).
23. Kaplan, J. *et al.* A key role for mitochondrial gatekeeper pyruvate dehydrogenase in oncogene-induced senescence. *Nature* **498**, 109–112 (2013).
24. Kondoh, H. *et al.* Glycolytic enzymes can modulate cellular life span. *Cancer Res.* **65**, 177–185 (2005).
25. Dörr, J. R. *et al.* Synthetic lethal metabolic targeting of cellular senescence in cancer therapy. *Nature* **501**, 421–425 (2013).

References 20, 23, 25 revealed that metabolic mechanisms are causally implicated in the induction and maintenance of the senescent state.

26. Shamma, A. *et al.* Rb Regulates DNA damage response and cellular senescence through E2F-dependent suppression of N-ras isoprenylation. *Cancer Cell* **15**, 255–269 (2009).
27. Ben-Porath, I. & Weinberg, R. A. The signals and pathways activating cellular senescence. *Int. J. Biochem. Cell Biol.* **37**, 961–976 (2005).
28. Moiseeva, O., Mallette, F. A., Mukhopadhyay, U. K., Moores, A. & Ferbeyre, G. DNA damage signaling and p53-dependent senescence after prolonged beta-interferon stimulation. *Mol. Biol. Cell* **17**, 1583–1592 (2006).

29. Romanov, V. S. *et al.* p21(Waf1) is required for cellular senescence but not for cell cycle arrest induced by the HDAC inhibitor sodium butyrate. *Cell Cycle* **9**, 3945–3955 (2010).
30. Lapak, K. M. & Burd, C. E. The molecular balancing act of p16^{INK4a} in cancer and aging. *Mol. Cancer Res.* **167**–183 (2014).
31. Hein, N. *et al.* in *Senescence* (ed. Nagata, T.) Ch. 9, 171–208 (2012).
32. Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
33. Schmidt, S. *et al.* The centrosome and mitotic spindle apparatus in cancer and senescence. *Cell Cycle* **9**, 4469–4473 (2010).
34. Freund, A., Patil, C. K. & Campisi, J. p38MAPK is a novel DNA damage response-independent regulator of the senescence-associated secretory phenotype. *EMBO J.* **30**, 1536–1548 (2011).
35. Passos, J. F. *et al.* Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Mol. Syst. Biol.* **6**, 347 (2010).
36. Baker, D. J., Weaver, R. L. & van Deursen, J. M. p21 both attenuates and drives senescence and aging in BubR1 progeroid mice. *Cell Rep.* **3**, 1164–1174 (2013).
37. Grove, G. L. & Cristofalo, V. J. Characterization of the cell cycle of cultured human diploid cells: effects of aging and hydrocortisone. *J. Cell. Physiol.* **90**, 415–422 (1977).
38. Choudhury, A. R. *et al.* Cdkn1a deletion improves stem cell function and lifespan of mice with dysfunctional telomeres without accelerating cancer formation. *Nature Genet.* **39**, 99–105 (2007).
39. Siegel, J. J. & Amon, A. New insights into the troubles of aneuploidy. *Annu. Rev. Cell Dev. Biol.* **28**, 189–214 (2012).
40. Vredevelde, L. C. *et al.* Abrogation of BRAF^{V600E}-induced senescence by PI3K pathway activation contributes to melanomagenesis. *Genes Dev.* **26**, 1055–1069 (2012).
41. Lapasset, L. *et al.* Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. *Genes Dev.* **25**, 2248–2253 (2011).
42. De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247–256 (2013).
43. Wang, J. *et al.* Inhibition of activated pericentromeric SINE/Alu repeat transcription in senescent human adult stem cells reinstates self-renewal. *Cell Cycle* **10**, 3016–3030 (2011).
44. Ivanov, A. *et al.* Lysosome-mediated processing of chromatin in senescence. *J. Cell Biol.* **202**, 129–143 (2013).
45. Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science* **336**, 1440–1444 (2012).
46. Freund, A., Laberge, R. M., Demaria, M. & Campisi, J. Lamin B1 loss is a senescence-associated biomarker. *Mol. Biol. Cell* **23**, 2066–2075 (2012).
47. Shimi, T. *et al.* The role of nuclear lamin B1 in cell proliferation and senescence. *Genes Dev.* **25**, 2579–2593 (2011).
48. Shah, P. P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* **27**, 1787–1799 (2013).
49. Funayama, R., Saito, M., Tanobe, H. & Ishikawa, F. Loss of linker histone H1 in cellular senescence. *J. Cell Biol.* **175**, 869–880 (2006).
50. Narita, M. *et al.* A novel role for high-mobility group A proteins in cellular senescence and heterochromatin formation. *Cell* **126**, 503–514 (2006).
51. Zhang, R., Chen, W. & Adams, P. D. Molecular dissection of formation of senescence-associated heterochromatin foci. *Mol. Cell Biol.* **27**, 2343–2358 (2007).
52. Swanson, E. C., Manning, B., Zhang, H. & Lawrence, J. B. Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence. *J. Cell Biol.* **929**–942 (2013).
53. Shelton, D. N., Chang, E., Whittier, P. S., Choi, D. & Funk, W. D. Microarray analysis of replicative senescence. *Curr. Biol.* **9**, 939–945 (1999).
54. Zhang, H., Pan, K. H. & Cohen, S. N. Senescence-specific gene expression fingerprints reveal cell-type-dependent physical clustering of up-regulated chromosomal loci. *Proc. Natl Acad. Sci. USA* **100**, 3251–3256 (2003).
55. Coppé, J. P. *et al.* Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLoS Biol.* **6**, e301 (2008).
56. Rodier, F. *et al.* Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nature Cell Biol.* **11**, 973–979 (2009); erratum **11**, 1272 (2009).
57. Kuilman, T. & Peeper, D. S. Senescence-messaging secretome: SMS-ing cellular stress. *Nature Rev. Cancer* **9**, 81–94 (2009).
58. Coppé, J. P. *et al.* Tumor suppressor and aging biomarker p16^{INK4a} induces cellular senescence without the associated inflammatory secretory phenotype. *J. Biol. Chem.* **286**, 36396–36403 (2011).
59. Coppé, J. P., Kauser, K., Campisi, J. & Beausejour, C. M. Secretion of vascular endothelial growth factor by primary human fibroblasts at senescence. *J. Biol. Chem.* **281**, 29568–29574 (2006).
60. Freund, A., Orjalo, A. V., Desprez, P. Y. & Campisi, J. Inflammatory networks during cellular senescence: causes and consequences. *Trends Mol. Med.* **16**, 238–246 (2010).
61. Laberge, R. M., Awad, P., Campisi, J. & Desprez, P. Y. Epithelial-mesenchymal transition induced by senescent fibroblasts. *Cancer Microenviron.* **5**, 39–44 (2012).
62. Binet, R. *et al.* WNT16B is a new marker of cellular senescence that regulates p53 activity and the phosphoinositide 3-kinase/AKT pathway. *Cancer Res.* **69**, 9183–9191 (2009).
63. Kuilman, T. *et al.* Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* **133**, 1019–1031 (2008).

64. Yang, G. *et al.* The chemokine growth-regulated oncogene 1 (Gro-1) links RAS signaling to the senescence of stromal fibroblasts and ovarian tumorigenesis. *Proc. Natl Acad. Sci. USA* **103**, 16472–16477 (2006).
65. Benz, G., Holzel, D. & Schmoekel, C. Inflammatory cellular infiltrates in melanocytic nevi. *Am. J. Dermatopathol.* **13**, 538–542 (1991).
66. Finkel, T., Serrano, M. & Blasco, M. A. The common biology of cancer and ageing. *Nature* **448**, 767–774 (2007).
67. Matheu, A. *et al.* Delayed ageing through damage protection by the Arf/p53 pathway. *Nature* **448**, 375–379 (2007).
68. Baker, D. J. *et al.* Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. *Nature Cell Biol.* **15**, 96–102 (2013).
69. Jun, J. I. & Lau, L. F. Cellular senescence controls fibrosis in wound healing. *Aging* **2**, 627–631 (2010).
70. Wang, J., Geiger, H. & Rudolph, K. L. Immunoaging induced by hematopoietic stem cell aging. *Curr. Opin. Immunol.* **23**, 532–536 (2011).
71. Nikolich-Zugich, J. Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. *Nature Rev. Immunol.* **8**, 512–522 (2008).
72. Roninson, I. B. Tumor cell senescence in cancer treatment. *Cancer Res.* **63**, 2705–2715 (2003).
73. Le, O. N. *et al.* Ionizing radiation-induced long-term expression of senescence markers in mice is independent of p53 and immune status. *Aging Cell* **9**, 398–409 (2010).
74. Allan, J. M. & Travis, L. B. Mechanisms of therapy-related carcinogenesis. *Nature Rev. Cancer* **5**, 943–955 (2005).
75. Jurk, D. *et al.* Postmitotic neurons develop a p21-dependent senescence-like phenotype driven by a DNA damage response. *Aging Cell* **11**, 996–1004 (2012).
- A study showing that post-mitotic cells can obtain several key characteristics of senescent cells.**
76. Minamino, T. *et al.* A crucial role for adipose tissue p53 in the regulation of insulin resistance. *Nature Med.* **15**, 1082–1087 (2009).
77. Herbig, U., Ferreira, M., Condel, L., Carey, D. & Sedivy, J. M. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
78. Lawless, C. *et al.* Quantitative assessment of markers for cell senescence. *Exp. Gerontol.* **45**, 772–778 (2010).
79. Wang, C. *et al.* DNA damage response and cellular senescence in tissues of aging mice. *Aging Cell* **8**, 311–323 (2009).
80. Krishnamurthy, J. *et al.* p16^{INK4a} induces an age-dependent decline in islet regenerative potential. *Nature* **443**, 453–457 (2006).
81. Jeyapalan, J. C., Ferreira, M., Sedivy, J. M. & Herbig, U. Accumulation of senescent cells in mitotic tissue of aging primates. *Mech. Ageing Dev.* **128**, 36–44 (2007).
82. Naylor, R. M., Baker, D. J. & van Deursen, J. M. Senescent cells: a novel therapeutic target for aging and age-related diseases. *Clin. Pharmacol. Ther.* **93**, 105–116 (2013).
83. Sherr, C. J. The Pezcoller lecture: cancer cell cycles revisited. *Cancer Res.* **60**, 3689–3695 (2000).
84. Campisi, J. Cellular senescence: putting the paradoxes in perspective. *Curr. Opin. Genet. Dev.* **21**, 107–112 (2011).
85. Coppé, J. P., Desprez, P. Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu. Rev. Pathol.* **5**, 99–118 (2010).
86. Rodier, F. & Campisi, J. Four faces of cellular senescence. *J. Cell Biol.* **192**, 547–556 (2011).
87. Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807–810 (2007).
88. Krtolica, A. *et al.* G α regulates human embryonic stem cell self-renewal or adoption of a neuronal fate. *Differentiation* **81**, 222–232 (2011).
89. Pricola, K. L., Kuhn, N. Z., Haleem-Smith, H., Song, Y. & Tuan, R. S. Interleukin-6 maintains bone marrow-derived mesenchymal stem cell stemness by an ERK1/2-dependent mechanism. *J. Cell. Biochem.* **108**, 577–588 (2009).
90. Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760–764 (2005).
91. Parrinello, S., Coppe, J. P., Krtolica, A. & Campisi, J. Stromal-epithelial interactions in aging and cancer: senescent fibroblasts alter epithelial cell differentiation. *J. Cell Sci.* **118**, 485–496 (2005).
92. Acosta, J. C. *et al.* A complex secretory program orchestrated by the inflammasome controls paracrine senescence. *Nature Cell Biol.* **15**, 978–990 (2013).
93. Nelson, G. *et al.* A senescent cell bystander effect: senescence-induced senescence. *Aging Cell* **11**, 345–349 (2012).
- References 92, 94 show that senescent cells can induce senescence in neighbouring cells through paracrine mechanisms.**
94. Faggioli, F., Wang, T., Vijg, J. & Montagna, C. Chromosome-specific accumulation of aneuploidy in the aging mouse brain. *Hum. Mol. Genet.* **21**, 5246–5253 (2012).
95. Garinis, G. A., van der Horst, G. T., Vijg, J. & Hoeijmakers, J. H. DNA damage and ageing: new-age ideas for an age-old problem. *Nature Cell Biol.* **10**, 1241–1247 (2008).
96. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
97. Kang, T. W. *et al.* Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479**, 547–551 (2011).
98. Krishnamurthy, J. *et al.* Ink4a/Arf expression is a biomarker of aging. *J. Clin. Invest.* **114**, 1299–1307 (2004).
99. Bhaumik, D. *et al.* MicroRNAs miR-146a/b negatively modulate the senescence-associated inflammatory mediators IL-6 and IL-8. *Aging* **1**, 402–411 (2009).

Acknowledgements I thank J. Campisi, J. Kirkland, R. Naylor, B. Childs, D. Baker, R. Urrutia, M. McNiven and R. Bram for helpful discussions and comments on the manuscript. I apologize to those whom I was unable to reference owing to space limitations. This work was supported by grants from the Paul Glenn Foundation and the National Institutes of Health (R01CA96985, R01CA166347 and AG41122-01P2).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.v.D. (vandeursen.jan@mayo.edu).

Genome–defence small RNAs exapted for epigenetic mating–type inheritance

Deepankar Pratap Singh^{1,2}, Baptiste Saudemont^{1,2†}, Gérard Guglielmi¹, Olivier Arnaiz³, Jean–François Gout^{4†}, Malgorzata Prajer⁵, Alexey Potekhin⁶, Ewa Przybós⁵, Anne Aubusson–Fleury³, Simran Bhullar¹, Khaled Bouhouche^{1†}, Maoussi Lhuillier–Akakpo^{2,7}, Véronique Tanty¹, Corinne Blugeon¹, Adriana Alberti⁸, Karine Labadie⁸, Jean–Marc Aury⁸, Linda Sperling³, Sandra Duharcourt⁷ & Eric Meyer¹

In the ciliate *Paramecium*, transposable elements and their single-copy remnants are deleted during the development of somatic macronuclei from germline micronuclei, at each sexual generation. Deletions are targeted by scnRNAs, small RNAs produced from the germ line during meiosis that first scan the maternal macronuclear genome to identify missing sequences, and then allow the zygotic macronucleus to reproduce the same deletions. Here we show that this process accounts for the maternal inheritance of mating types in *Paramecium tetraurelia*, a long-standing problem in epigenetics. Mating type E depends on expression of the transmembrane protein *mtA*, and the default type O is determined during development by scnRNA-dependent excision of the *mtA* promoter. In the sibling species *Paramecium septaurelia*, mating type O is determined by coding-sequence deletions in a different gene, *mtB*, which is specifically required for *mtA* expression. These independently evolved mechanisms suggest frequent exaptation of the scnRNA pathway to regulate cellular genes and mediate transgenerational epigenetic inheritance of essential phenotypic polymorphisms.

Ciliates are complex unicellular eukaryotes that use different types of nuclei within the same cytoplasm to separate germline and somatic functions¹. The diploid micronuclei (MICs) undergo meiosis to provide gametic nuclei during sexual events, but their genome is not expressed. Genes are expressed from the polyploid macronucleus (MAC), which is not transmitted across sexual generations. After meiosis and fertilization, the parental MAC is lost and replaced by a new one that develops from a mitotic copy of the zygotic nucleus. Macronuclear development involves extensive rearrangements of the germline genome, including the elimination of virtually all transposable elements and other repeats². Furthermore, in *P. tetraurelia* ~45,000 short, single-copy internal eliminated sequences (IESs) are precisely excised from coding and non-coding sequences^{3,4} by the domesticated transposase Pgm⁵. IESs are invariably flanked by two 5′-TA-3′ dinucleotides which recombine into one after excision⁴. A short consensus adjacent to the TAs (5′-TAYAGYNR-3′) is reminiscent of the ends of Tc1/mariner elements^{4,6}, and a recent study provided support for the hypothesis that IESs are degenerate remnants of ancient transposable element insertions^{3,7}.

All intragenic IESs identified so far must be excised to reconstitute functional genes in the MAC. However, the poorly conserved IES end consensus is not sufficient to specify the excision pattern genome-wide⁴, and a specific class of small RNAs is required to identify some IESs on the basis of their absence from the parental MAC^{8–10}. In the current ‘genome scanning’ model, scnRNAs are produced from most of the germline genome during MIC meiosis^{11,12} and are then filtered by pairing interactions with nascent transcripts in the parental MAC¹³, which acts as a sponge to remove matching scnRNAs from the active pool. Those that cannot find a match remain free to target homologous sequences in the zygotic MAC when it develops, thereby recruiting the IES excision machinery. This

RNA-mediated genomic subtraction thus reproduces the deletions observed in the parental MAC and can account for epigenetic inheritance of alternative rearrangement patterns, such as retention of a given IES in the MAC^{14,15}, or deletion of a given gene¹⁶, across sexual generations.

In *Paramecium*, conjugation (the reciprocal fertilization of cells of opposite mating types) does not allow any significant exchange of cytoplasm between the mates, so that the pools of scnRNAs produced during meiosis are independently sorted in each cell. After fertilization, each developing MAC will thus reproduce the particular rearrangements present in the old MAC of its own cytoplasmic parent. This mechanism might underlie the maternal (cytoplasmic) inheritance of mating types in *P. tetraurelia*, one of the earliest cases of transgenerational epigenetic inheritance in any eukaryote.

Although mating types were discovered in 1937 (ref. 17), so far the only available test relies on the ability of type O (odd) to agglutinate with type E (even) when vegetative cells become sexually reactive, a physiological state induced by mild starvation. Agglutination is a prerequisite for conjugation and occurs on contact through adhesion of ciliary membranes. Mutational analyses showed that several genes are specifically required for expression of type E^{18–20}; mutations in these genes result in a constitutive O phenotype, which thus seems to be a default state. In *P. tetraurelia*, mating types are not genetically determined in the MIC²¹. Each new MAC becomes determined for one type during its development, and remains the same throughout vegetative growth of the derived clone. The O/E alternative is not random, but maternally inherited; experiments showed that mating-type determination in the developing zygotic MAC is controlled through the cytoplasm²² by the maternal MAC^{23,24}. A pleiotropic mutation enforcing constitutive determination for type E was later found to impair a limited subset of genome rearrangements²⁵, suggesting that type O is normally

¹Ecole Normale Supérieure, Institut de Biologie de l’ENS, IBENS; Inserm, U1024; CNRS, UMR 8197 Paris F-75005, France. ²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex 05, France. ³CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198, and Université Paris-Sud, Département de Biologie, Orsay F-91405, France. ⁴CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, 43 boulevard du 11 Novembre 1918, Villeurbanne F-69622, France. ⁵Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Ślaskowska 17, 31-016 Krakow, Poland. ⁶Department of Microbiology, Faculty of Biology, St Petersburg State University, Saint Petersburg 199034, Russia. ⁷Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris F-75205, France. ⁸Commissariat à l’Energie Atomique (CEA), Institut de Génétique (IG), Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France. [†]Present addresses: Laboratoire de Biochimie, Unité Mixte de Recherche 8231, École Supérieure de Physique et de Chimie Industrielles, 75231 Paris, France (B.S.); Department of Biology, Indiana University, Bloomington, Indiana 47405, USA (J.F.G.); INRA, UMR 1061 Unité de Génétique Moléculaire Animale, Université de Limoges, IFR 145, Faculté des Sciences et Techniques, 87060 Limoges, France (K.B.).

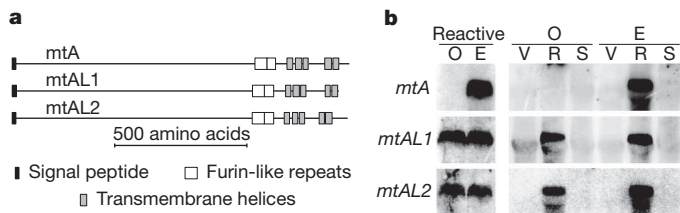


Figure 1 | Structure and expression of *mtA* and related proteins.

a, Recognizable protein features. **b**, Northern blot analysis of expression of *mtA*, *mtAL1* (GSPATG00025159001) and *mtAL2* (GSPATG00002922001). The panels on the left compare total RNA samples from sexually reactive cultures of the two mating types (reactive, O and E). In the panels on the right, RNA samples were extracted from: V, vegetative, exponentially growing cells; R, mildly starved and sexually reactive; S, starved for 48 h and no longer reactive. The three mRNAs migrate just above the large rRNA, as can be seen from the pattern of background hybridization.

determined by the rearrangement of one of the genes required for E expression²⁶.

mtA rearrangements determine mating type

To identify the putative mating-type-determining gene, we used a whole-genome microarray to compare the transcriptomes of sexually reactive cells of both mating types (not shown). The gene found to have the largest E/O expression ratio, GSPATG00017533001, encodes a 1,275-amino-acid protein with a signal peptide and cysteine-rich furin-like repeats followed by 5 transmembrane segments at the carboxy terminus (Fig. 1a and Supplementary Data 1a); the amino-terminal part is predicted to be outside the plasma membrane. This gene was later shown to be *mtA* (ref. 19) (see below). Northern blot analyses confirmed that it is expressed only in E cells and further showed that expression is limited to sexual reactivity, as transcripts could not be detected in exponentially growing cells or in over-starved cells (Fig. 1b). Genes encoding structurally similar proteins were also found to be specifically expressed in sexually reactive cells, but in both mating types (Fig. 1b). RNA interference (RNAi)-mediated silencing of *mtA* in E cells resulted in the default O phenotype during sexual reactivity (as determined by their capacity to agglutinate with E but not with O tester lines), indicating that *mtA* is required for E expression (Supplementary Table 1).

To understand the molecular basis for mating-type-E-specific expression, we sequenced the *mtA* gene from the MACs of O and E cells, and the unarranged MIC version. The MIC gene is interrupted by four IESs that are excised in both MAC types; however, a 195-base-pair (bp) segment containing the transcription start site and the first 26 bp of the coding sequence was found to be excised as an IES in mating type O MACs, but retained in E MACs (Fig. 2a–c and Supplementary Data 1a). This segment contains the *mtA* promoter, as indicated by microinjection of different constructs into the MACs of O cells (Extended Data Fig. 1a–c): its presence upstream of the coding sequence was sufficient for transformed clones to express mating type E instead of mating type O. An *mtA*–GFP fusion protein was detected in cilia of the anteroventral surface, although an excess of GFP fluorescence was seen in cytoplasmic structures, probably the ER (Fig. 2d, e). The fusion protein could be detected on cilia with anti-GFP antibodies in fixed, non-permeabilized cells (Fig. 2f), confirming that its N-terminal part is exposed outside of the ciliary membrane.

When *mtA*-transformed clones were taken through autogamy (a self-fertilization sexual process), the induced type E was robustly transmitted to progeny (Extended Data Fig. 1b). As observed for maternally controlled IESs^{14,15}, a plasmid containing only the 195-bp segment was sufficient to inhibit excision of the homologous sequence during development of zygotic MACs, causing the progeny of transformed O cells to switch to E (Extended Data Fig. 1d). Retention of the *mtA* promoter thus recapitulates both aspects of developmental determination for E, namely the capacity to express mating-type E during the vegetative phase and the capacity, after sexual events, to direct determination of new MACs for the same type.

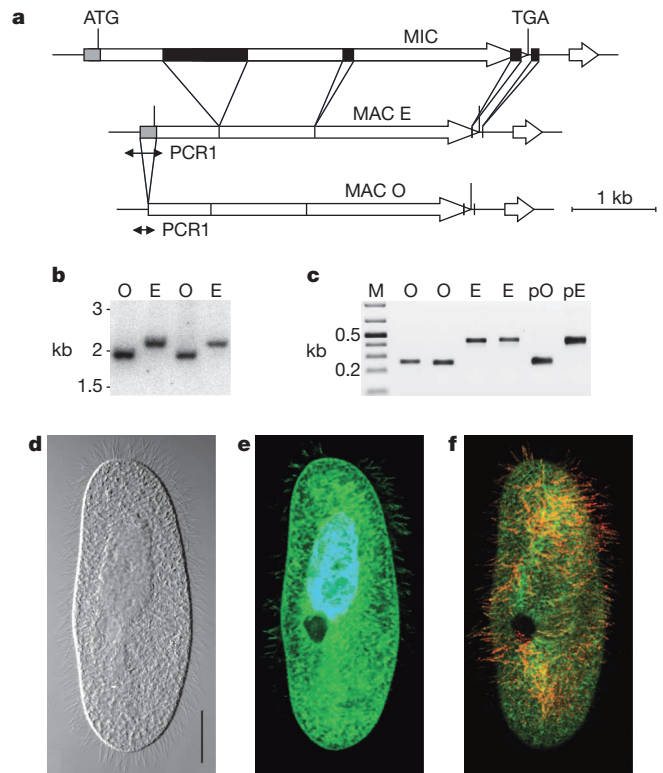


Figure 2 | Structure of *mtA* in the MIC and in the MACs of E and O cells.

a, Black boxes are IESs; the 195-bp segment (grey box) is excised only in O MACs. A small unannotated gene downstream of *mtA* is conserved in other species. **b**, Southern blot of EcoRI-digested total DNA, hybridized with an *mtA* probe (PCR5) revealing fragments of 1,945 and 2,140 bp in O and E clones. **c**, PCR amplification of the *mtA* 5' end (PCR1) on independent pools of O and E clones (266 and 461 bp, respectively). pO and pE, control PCRs on cloned MAC versions; M, size markers. **d–f**, Localization of an *mtA*–GFP fusion protein, with GFP inserted between furin-like repeats and transmembrane segments. **d**, DIC image. Scale bar, 20 μ m. **e**, Intracellular confocal optical section. Some GFP fluorescence (green) is detected in cilia of the anterior (top), but not posterior, part of the cell. The MAC is stained with Hoechst (blue). **f**, Confocal image of the ventral cell surface, showing the opening of the oral apparatus. The fixed, non-permeabilized cell was labelled with anti-GFP and secondary antibodies (red).

Expression and inheritance uncoupled

Previous mutational analyses showed that the expression of mating types during sexual reactivity can be uncoupled from the mechanism that ensures their transgenerational inheritance^{19–21,23}. The recessive mutations *mtA*^O, *mtB*^O and *mtC*^O all preclude expression of type E and restrict cells to the default O phenotype, but they do not affect the maternal inheritance of mating-type determination: mutant homozygotes formed in an E cytoplasmic lineage, despite expressing O, will keep the memory of E determination for an indefinite number of sexual generations, as shown by the fact that their progeny switches back to E expression when the wild-type allele is reintroduced by conjugation¹⁹ (Extended Data Fig. 2a–d).

To determine whether this memory is attributable to retention of the *mtA* promoter, we crossed each of the mutants to the wild type. Autogamy of F₁ heterozygotes of mating-type E yielded homozygous F₂ progeny with a 1:1 ratio of E- and O-expressing clones, reflecting the Mendelian segregation of wild-type and mutant alleles. Both types of clones retained the *mtA* promoter (Extended Data Fig. 2e). That the O-expressing mutant clones could still transmit E determination was verified by crossing them again to wild-type E cells: the second-round F₁ heterozygotes derived from both parents always expressed E and retained the *mtA* promoter (Extended Data Fig. 2f). Autogamy of these F₁ heterozygotes again resulted in the expected Mendelian segregation of E- and O-expressing clones among second-round F₂ homozygotes (Supplementary Table 2). Thus, independently

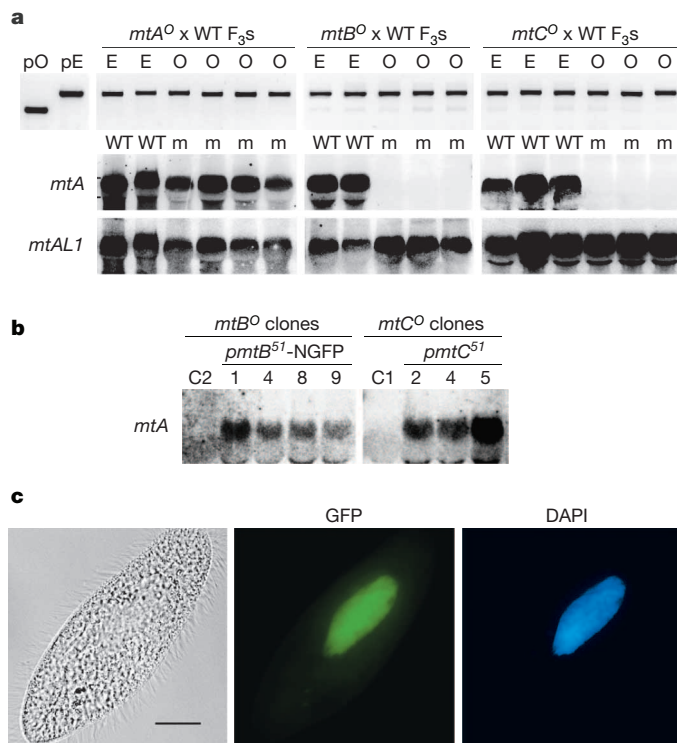


Figure 3 | Molecular analysis of expression mutants. **a**, *mtA* expression in E-determined *mtA^O*, *mtB^O* and *mtC^O* homozygotes. The top panel shows a PCR analysis (PCR1) of E- and O-expressing F₃ lines from the second round of backcross of each mutant to wild-type E cells. pO and pE, control PCRs on cloned MAC versions of *mtA*. All clones retained the *mtA* promoter. The wild-type (WT) or mutant (m) genotypes were confirmed by sequencing. Middle and bottom panels show Northern blots of total RNA samples from sexually reactive cultures of each F₃ line, hybridized with *mtA* or *mtAL1* probes. **b**, Northern blot analysis of *mtA* expression in sexually reactive cultures of E-determined *mtB^O* and *mtC^O* clones transformed with plasmids *pmtB⁵¹-NGFP* and *pmtC⁵¹*, respectively (see Extended Data Fig. 3). Transformed clones expressed mating type E, whereas uninjected controls (C2 and C1) expressed mating type O. **c**, Fluorescence microscopy of an *mtB^O* cell transformed with *pmtB⁵¹-NGFP*, expressing mating type E (GFP and DAPI filters). Scale bar, 20 μ m.

of the mating type being expressed, maternal inheritance of *mtA* promoter retention may indeed underlie the epigenetic memory of E determination.

To understand how mutant clones can express mating type O despite retaining the *mtA* promoter, we tested *mtA* expression in E- and O-expressing lines, using F₃ populations obtained by an additional autogamy of second-round F₂ homozygotes. In the case of the *mtA^O* mutation, Northern blots showed that *mtA* (GSPATG00017533001) transcripts were produced in sexually reactive cultures of O-expressing lines (Fig. 3a), prompting us to re-sequence the gene in mutant clones. A substitution was found to change Arg codon 751 to a stop codon, and the mutation co-segregated with mating-type O expression among second-round F₂ homozygotes (Supplementary Table 2a). This revealed that GSPATG00017533001 is *mtA* and confirmed that the encoded protein is required for expression of mating type E. The mutation lies at a distance from the *mtA* promoter and apparently does not affect the regulation of its excision, explaining how O-expressing mutant homozygotes can transmit either O or E determination.

In contrast to *mtA^O* lines, *mtB^O* and *mtC^O* O-expressing lines did not produce *mtA* mRNA on sexual reactivity (Fig. 3a). The mutations were identified by whole-genome sequencing, and found to be substitutions in GSPATG00026812001 and GSPATG00009074001, respectively (Supplementary Data 1b, c). Correct identification of the mutations was confirmed by their co-segregation with O expression among F₂ lines (Supplementary Table 2c, d), and by transformation of the MACs of E-determined, O-expressing mutants with the wild-type alleles, which resulted

in *mtA* transcription and E expression in both cases (Fig. 3b and Extended Data Fig. 3). Thus, both *mtB* and *mtC* seem to be transcription factors required for *mtA* expression in E-determined cells. The *mtB* and *mtC* genes are constitutively expressed at low levels, and both of the encoded proteins are predicted to be nuclear. *mtC* is a 138-amino-acid protein containing a C2H2 zinc finger, the structure of which is probably affected by the *mtC^O* mutation (Supplementary Data 2). Although the 310-amino-acid *mtB* protein does not contain any recognizable domain, an *mtB*-GFP fusion protein was found to localize to the MAC (Fig. 3c). Deep sequencing of the transcriptomes of *mtB^O* and *mtC^O* mutants further showed that *mtA* is the only gene that requires both factors for expression during sexual reactivity; importantly, it is the only gene found to be downregulated in the *mtB^O* mutant (Supplementary Table 3).

mtA promoter excision is regulated by scnRNAs

Excision of the *mtA* promoter in mating-type O cells occurs between two 5'-TA-3' dinucleotides within a reasonable IES end consensus (Supplementary Data 1a), suggesting that the general IES excision machinery is involved. RNAi-mediated depletion of the Pgm endonuclease⁵ during autogamy of O cells indeed impaired excision and resulted in accumulation of unexcised copies in the developing new MAC (Fig. 4a). So did depletion of proteins known to be involved in scnRNA biogenesis or action, including the Dicer-like Dcl2 and Dcl3 (ref. 12), the Piwi-like Ptiwi01 and Ptiwi09 (ref. 11), and the Nowa1 and Nowa2 RNA-binding proteins²⁷, indicating that the scnRNA pathway is required to target excision in O cells. Depletion of each pair of proteins causes massive retention of IESs genome-wide, resulting in non-functional new MACs and post-autogamous lethality. In conditions of partial depletion, obtained by silencing only one of the two Piwi (or the two Dicer-like) genes, the progeny is usually viable. Analysis of individual clones showed that they frequently retained the *mtA* promoter in the new MAC, and in all cases this correlated with a switch to mating type E (Fig. 4b).

No other IES is known to be affected in these conditions; in this regard, excision of the *mtA* promoter behaves like MAC deletions of cellular genes, which were shown to be more sensitive to partial impairment of the scnRNA pathway^{11,12}. Whereas other IESs are intervening sequences

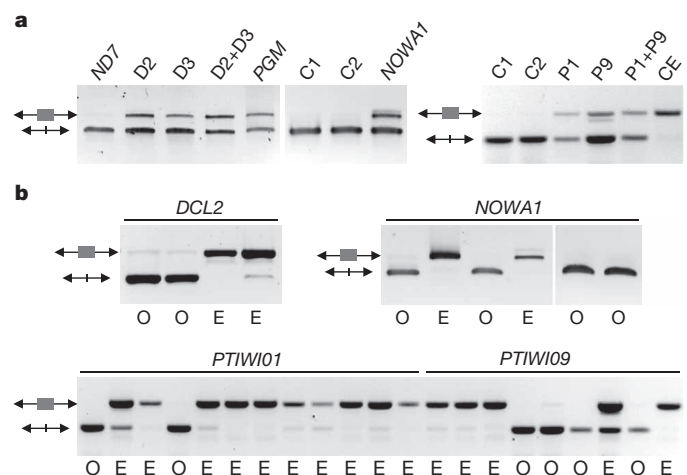


Figure 4 | Genes required for excision of the *mtA* promoter in O cells. **a**, PCR analysis (PCR5) of *mtA* promoter retention in mass progenies of O clones after RNAi-mediated silencing of the indicated genes. Total DNA samples were prepared from starved post-autogamous cells when the new MACs were clearly visible. Because the parental MAC is still present at this stage, the promoter-excised version is amplified in all cases; the promoter-retaining fragment can be detected only if it accumulates in zygotic MACs. ND7, unrelated-gene negative control; D2, *DCL2*; D3, *DCL3*; C1, no-silencing negative control; C2, empty-vector RNAi control; P1, *PTIWI01*; P9, *PTIWI09*; CE, control PCR5 on E cells. **b**, PCR analysis (PCR5) and mating-type tests (below) of individual viable post-autogamous clones from non-lethal silencing conditions (including partial silencing of *NOWA1*).

that must be excised to reconstitute functional genes, the 195-bp segment is a functional part of the *mtA* gene. The homologous segment is not excised in *mtA* orthologues from other *P. aurelia* species (see below), indicating that excision of a bona fide part of the MAC genome evolved as a derived character in *P. tetraurelia*, as random mutations happened to create sites suitable for Pgm-mediated excision.

Deep sequencing of small RNAs from an early conjugation time point confirmed that scnRNAs are produced from both strands of the *mtA* promoter, as they are from the rest of the germline genome (Extended Data Fig. 4). On average, scnRNA coverage was ~2-fold higher for IESs than for other genome features, as independently found in another study²⁸. Although we cannot exclude that this results from precocious degradation of MAC-matching scnRNAs, an intrinsically more abundant production of scnRNAs from IESs could explain why IES excision is less sensitive to partial impairment of the scnRNA pathway than gene deletion. Northern blot analyses of autogamy time courses showed that *mtA*-promoter scnRNAs are produced in similar amounts during meiosis of O and E cells, and failed to detect any difference in the timing of their disappearance (Extended Data Fig. 5). Thus, *mtA*-promoter scnRNAs may be inactivated in E cells by some other mechanism than the active degradation proposed in *Tetrahymena thermophila*^{29–31}, for instance by sequestration in the maternal MAC. Whatever the precise mechanism, one prediction of the model is that RNAi-mediated destruction of maternal transcripts of the *mtA* promoter during autogamy of E cells will prevent inactivation of homologous scnRNAs, licensing them to target excision in the zygotic MAC¹³. Indeed, feeding E cells with double-stranded RNA homologous to the *mtA* promoter before autogamy induced its precise excision in the new MAC and resulted in O progeny (Supplementary Table 4). The induced type O was thereafter inherited for at least three sexual generations (not shown).

A different switch in *P. septaurelia*

P. tetraurelia belongs to a group of 15 sibling species that are morphologically indistinguishable but sexually incompatible^{1,32}. All have homologous O and E mating types, as shown by cross-agglutination between the O types of some species and the E types of others. However, maternal inheritance of mating types is observed in only half of these species. In others, mating types are randomly determined during MAC development, without any influence of the maternal MAC. The distribution of these systems in the phylogenetic tree of *aurelia* species suggests multiple changes during evolution of the complex^{33,34} (Extended Data Fig. 6). We first sequenced *mtA*, *mtB* and *mtC* orthologues in two strains of *P. octaurelia*, a maternal-inheritance species closely related to *P. tetraurelia*, with which it can form viable (though sterile) F₁ hybrids³⁵. We found the *mtA* promoter to be excised in O clones, with the same boundaries as in *P. tetraurelia*, but not in E clones (Extended Data Fig. 7 and Supplementary Data 1a). This excision system thus probably evolved in the common ancestor of the two species.

We then examined *P. septaurelia*, another maternal-inheritance species that groups with random-determination species. Sequencing of the *mtA*, *mtB* and *mtC* orthologues in the wild-type strain 227 showed that the *mtA*²²⁷ promoter was not excised in the MACs of O cells (Supplementary Data 1a). Instead, MAC copies of the *mtB*²²⁷ gene contained either of two alternative deletions of coding-sequence segments between IES-like boundaries (Fig. 5a and Supplementary Data 1b), suggesting that the gene was rearranged into non-functional forms. Our stock of strain 227 contained only O cells, but complementation of their MACs by microinjection of the *P. tetraurelia* *pmtB*⁵¹ transgene made these cells phenotypically E and resulted in *mtA* transcription during sexual reactivity (Fig. 5b). Thus, mating type E is characterized by *mtB*-dependent *mtA* expression in *P. septaurelia*, as in *P. tetraurelia*.

The *mtB* gene belongs to a family of up to six paralogues in *P. aurelia* species, and sequence conservation indicates that all have been under purifying selection since before speciation (not shown). The excision events that inactivate *mtB*²²⁷ in O cells are not observed in any other paralogue or in *mtB* orthologues from other species, where the 5' TAs correspond

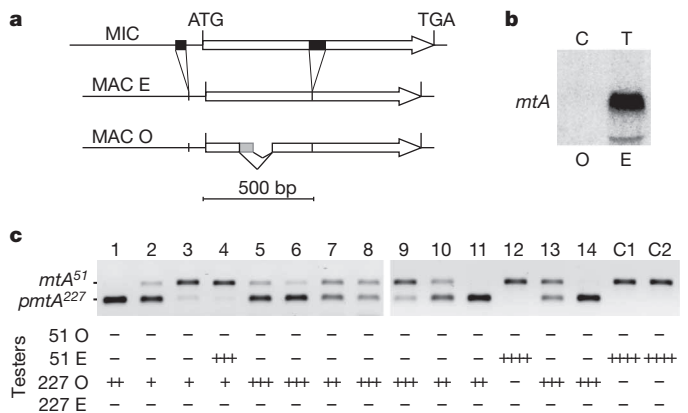


Figure 5 | *mtB* rearrangements determine mating types in *P. septaurelia*. **a**, MIC and MAC versions of *mtB*²²⁷. The MIC sequence contains two IESs (black boxes), conserved in *P. tetraurelia* strain 51. In the MAC of O cells of strain 227, the *mtB*²²⁷ coding sequence suffers two alternative deletions, each between TA dinucleotides within IES-end consensus sequences (TAYAG, see Supplementary Data 1b). **b**, Northern blot analysis of *mtA* expression in a sexually reactive, *pmtB*⁵¹-transformed O clone of strain 227 (T). C, control uninjected O clone. Tested mating types are indicated. **c**, Transformation of *P. tetraurelia* clones of mating type O with the *P. septaurelia* *mtA*²²⁷ gene. A duplex PCR was used to assess transgene copy numbers through the relative abundance of a 176-bp transgene-specific product (*pmtA*²²⁷, PCR12), compared to the 266-bp product from the promoterless endogenous *mtA* (*mtA*⁵¹, PCR1; undetectable in high-copy transformants). C1 and C2, uninjected control clones. Each clone was tested with *P. tetraurelia* (51) and *P. septaurelia* (227) O and E testers. Number of plus signs represents strength of mating reaction; minus indicates no agglutination.

to different nucleotides. As with *mtA* promoter excision in *P. tetraurelia*, this excision system apparently does not originate from any transposable element insertion but seems to have arisen in *P. septaurelia* after the chance appearance of good matches to the IES end consensus, which seems to be conserved among *P. aurelia* species³⁶.

Taking advantage of the evidence that the *mtB*⁵¹ protein from *P. tetraurelia* can activate the *mtA*²²⁷ promoter in *P. septaurelia*, we microinjected the *mtA*²²⁷ gene into the MACs of *P. tetraurelia* cells of mating type O, and tested the sexual preference of transformed clones with O and E testers of both species (Fig. 5c). Consistent with the reported lack of cross-agglutination between the complementary types of these two species¹, uninjected control clones reacted only with *P. tetraurelia* E testers. In contrast, transformed clones agglutinated only with *P. septaurelia* O testers, allowing interspecific conjugation to occur. Thus, *mtA*²²⁷ is sufficient for expression of the *P. septaurelia* E specificity, incidentally confirming that *mtA* localizes on the external side of ciliary membranes. The lack of reaction of transformed clones with *P. tetraurelia* O testers indicates that *mtA* interacts with an unknown O-specific receptor which also differs between species. Furthermore, transformed clones did not react with *P. tetraurelia* E testers, indicating that *mtA*²²⁷ expression, like *mtA*⁵¹ expression (Extended Data Fig. 1), functionally masks expression of the endogenous O-specific receptor. This would explain the default nature of the O type.

Strain 38 is a natural *P. septaurelia* isolate carrying a Mendelian mutation at a single locus (*mt*) which, in contrast to the *P. tetraurelia* mutations, affects both the expression and the epigenetic inheritance of mating types²⁰. This dual effect suggests that *mt* is the gene that controls mating types through an alternative rearrangement in that species (see Extended Data Fig. 8), providing an independent means to confirm the identification of the *P. septaurelia* switch. Sequencing of the *mtB*³⁸ allele revealed mutations that can explain both effects (Extended Data Fig. 8c and Supplementary Data 1b), and the analysis of a cross between strains 38 and 227 showed that *mt* is identical or closely linked to *mtB* (Supplementary Table 5). To confirm directly that *mtB* rearrangements are regulated by homology-dependent maternal effects, the full-length, functional MAC form

of the *mtB*²²⁷ allele was microinjected into the MACs of O cells of strain 227. Transformed clones expressed E and produced E post-autogamous progeny (Extended Data Fig. 9). We conclude that mating types are determined in *P. septaurelia* by maternally inherited alternative rearrangements of the *mtB* gene, a mechanism that must have evolved independently from that of *P. tetraurelia*.

Discussion

Seventy-five years after mating types were first described in *P. tetraurelia*, the expression of a single gene, *mtA*, was found to make the difference between the two types. *mtA* transcription in E cells was further shown to require the *mtB* and *mtC* gene products, and this E-specific pathway seems to be conserved in other *P. aurelia* species. Our results indicate that *mtA* is a ciliary transmembrane protein directly involved in the species-specific recognition of O cells. Although the nature of the O-specific receptor remains unknown, its expression or function appears to be masked by the *mtA* protein, given the default O phenotype observed in mutants. Expressing the *P. septaurelia* *mtA* orthologue in O cells of *P. tetraurelia* not only provided them with the *P. septaurelia* E specificity, but also blocked expression of the *P. tetraurelia* O specificity. In the related *P. caudatum*, an antigen involved in agglutination was shown to be produced in the cytoplasm of both O and E reactive cells, but to localize in the ciliary membrane only in O cells³⁷. In E cells of *P. aurelia* species, the *mtA* protein could inhibit the synthesis, processing or transport of the O-specific receptor. *mtA* is structurally similar to the proteins recently shown to be specific for each of the seven mating types of *T. thermophila*³⁸, but how these function in mating-type recognition remains to be determined.

Despite conservation of the *mtA* on/off switch as the E/O expression determinant, the mechanisms used to produce O clones differ among *P. aurelia* species. Excision of the *mtA* promoter is shared between the closely related *P. tetraurelia* and *P. octaurelia*, but in *P. septaurelia* *mtA* expression is prevented by deletions in *mtB* (Extended Data Fig. 10). This seems to be functionally equivalent, as *mtB* is highly specific for the *mtA* promoter in *P. tetraurelia*. The *P. tetraurelia* rearrangement involves the IES excision machinery and is regulated by the scnRNA pathway, explaining the maternal inheritance of mating types in this species, and probably in the other two as well. Ciliate scnRNAs resemble metazoan piRNAs in many respects, including their central roles in epigenetic reprogramming events that are critical for sexual reproduction³⁹, and both have been described as genomic 'immune systems' primarily involved in the control of transposable elements^{2,40–42}. One difference, however, is that metazoan piRNA systems keep a memory of 'non-self' sequences, based on integration into specific small-RNA-producing loci^{40,42,43}. The *Paramecium* system seems to work the opposite way, keeping a memory of 'self' sequences in the form of the maternal MAC, which is probed by genome-wide scnRNAs at each generation to identify 'non-self' molecules by subtraction.

As a consequence, the *Paramecium* genome-defence mechanism can easily be recruited for heritable regulation of bona fide cellular genes, an important question that has been raised in diverse eukaryotes^{42,44–46}. Indeed, the sequences being excised from the *mtA* and *mtB* genes differ from other IESs^{3,7} in that they are functional parts of cellular genes and apparently do not derive from the insertion of transposable elements; moreover, the *mtA* rearrangement behaves like deletions of cellular genes in its mechanistic requirements, which show subtle differences with IES excision^{11,12}. The independent evolution of the *P. tetraurelia* and *P. septaurelia* switches suggests that exaptation of the scnRNA pathway is a general mechanism for transgenerational epigenetic inheritance of differentiated states. In plants and mammals, epigenetic regulation of cellular genes often relies on transposon-induced epigenetic changes^{44,45}, as is the case of sex determination in melon⁴⁷. The capacity of the *Paramecium* mechanism to inactivate any single-copy gene makes it even more flexible, dispensing with the need for transposon insertion. In keeping with the low information content of the IES end consensus, accidental rearrangement errors are not rare genome-wide^{3,36,48}, and may occasionally provide a selective advantage in specific conditions; maternal

transmission to sexual progeny would thus allow continuous adaptation of the somatic (MAC) genome, independently of any Mendelian (MIC) mutation.

METHODS SUMMARY

Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*; the origins of other strains and species are given in the Methods section. Molecular biology experiments used standard procedures. *Paramecium*-specific methods (RNAi-mediated gene silencing by dsRNA feeding, DNA microinjection into the MAC) have been published and are referenced in the Methods section. For mating-type tests, testers were prepared from cell lines of known mating types by re-feeding ~1,000 autogamous cells in tubes for 2–3 divisions. The procedures used to test mass post-autogamous progenies or individual clones are presented in the Methods section. Bioinformatic analyses and statistical testing (microarray expression data, whole-genome transcriptome profiling by RNA-seq) are described in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 August 2013; accepted 11 April 2014.

Published online 7 May 2014.

1. Sonneborn, T. M. *Paramecium aurelia*. in *Handbook of Genetics* (ed. King, R. C.) 469–594 (Plenum, 1974).
2. Chalker, D. L. & Yao, M. C. DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet.* **45**, 227–246 (2011).
3. Arnaiz, O. et al. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.* **8**, e1002984 (2012).
4. Bétermier, M. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.* **155**, 399–408 (2004).
5. Baudry, C. et al. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* **23**, 2478–2483 (2009).
6. Klobutcher, L. A. & Herrick, G. Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic Acids Res.* **23**, 2006–2013 (1995).
7. Dubois, E. et al. Transposon invasion of the *Paramecium* germline genome countered by a domesticated PiggyBac transposase and the NHEJ pathway. *Int. J. Evol. Biol.* **2012**, 4361196 (2012).
8. Chalker, D. L., Meyer, E. & Mochizuki, K. Epigenetics of ciliates. *Cold Spring Harb. Perspect. Biol.* **5**, a017764 (2013).
9. Coyne, R. S., Lhuillier-Akakpo, M. & Duhaucourt, S. RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biol. Cell* **104**, 309–325 (2012).
10. Duhaucourt, S., Lepere, G. & Meyer, E. Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts. *Trends Genet.* **25**, 344–350 (2009).
11. Bouhouche, K., Gout, J. F., Kapusta, A., Bétermier, M. & Meyer, E. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* **39**, 4249–4264 (2011).
12. Lepere, G. et al. Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* **37**, 903–915 (2009).
13. Lepere, G., Bétermier, M., Meyer, E. & Duhaucourt, S. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* **22**, 1501–1512 (2008).
14. Duhaucourt, S., Butler, A. & Meyer, E. Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* **9**, 2065–2077 (1995).
15. Duhaucourt, S., Keller, A. M. & Meyer, E. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol.* **18**, 7075–7085 (1998).
16. Garnier, O., Serrano, V., Duhaucourt, S. & Meyer, E. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* **24**, 7370–7379 (2004).
17. Sonneborn, T. M. Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA* **23**, 378–385 (1937).
18. Butzel, H. M. Mating type mutations in variety 1 of *Paramecium aurelia*, and their bearing upon the problem of mating type determination. *Genetics* **40**, 321–330 (1955).
19. Byrne, B. C. Mutational analysis of mating type inheritance in Syngen 4 of *Paramecium aurelia*. *Genetics* **74**, 63–80 (1973).
20. Taub, S. R. The genetic control of mating type differentiation in *Paramecium*. *Genetics* **48**, 815–834 (1963).
21. Sonneborn, T. M. Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells. *Annu. Rev. Genet.* **11**, 349–367 (1977).
22. Sonneborn, T. M. Recent advances in the genetics of *Paramecium* and Euplotes. *Adv. Genet.* **1**, 263–358 (1947).

23. Nanney, D. L. Mating type inheritance at conjugation in variety 4 of *Paramecium aurelia*. *J. Protozool.* **4**, 89–95 (1957).
24. Sonneborn, T. M. Patterns of nucleo-cytoplasmic integration in *Paramecium*. *Caryologia* **6** (suppl.), 307–325 (1954).
25. Meyer, E. & Keller, A. M. A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in *Paramecium tetraurelia*. *Genetics* **143**, 191–202 (1996).
26. Meyer, E. & Garnier, O. Non-mendelian inheritance and homology-dependent effects in ciliates. *Adv. Genet.* **46**, 305–337 (2002).
27. Nowacki, M., Zagorski-Ostojka, W. & Meyer, E. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* **15**, 1616–1628 (2005).
28. Sandoval, P. Y., Swart, E. C., Arambasic, M. & Nowacki, M. Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell* **28**, 174–188 (2014).
29. Aronica, L. *et al.* Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.* **22**, 2228–2241 (2008).
30. Mochizuki, K. & Gorovsky, M. A. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.* **18**, 2068–2073 (2004).
31. Schoeberl, U. E., Kurth, H. M., Noto, T. & Mochizuki, K. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*. *Genes Dev.* **26**, 1729–1742 (2012).
32. Catania, F., Wurmser, F., Potekhin, A. A., Przybos, E. & Lynch, M. Genetic diversity in the *Paramecium aurelia* species complex. *Mol. Biol. Evol.* **26**, 421–431 (2009).
33. Hall, M. S. & Katz, L. A. On the nature of species: insights from *Paramecium* and other ciliates. *Genetica* **139**, 677–684 (2011).
34. Phadke, S. S. & Zufall, R. A. Rapid diversification of mating systems in ciliates. *Biol. J. Linn. Soc.* **98**, 187–197 (2009).
35. Haggard, B. W. Interspecies crosses in *Paramecium aurelia* (syngen 4 by syngen 8). *J. Protozool.* **21**, 152–159 (1974).
36. Catania, F., McGrath, C. L., Doak, T. G. & Lynch, M. Spliced DNA sequences in the *Paramecium* germline: their properties and evolutionary potential. *Genome Biol. Evol.* **5**, 1200–1211 (2013).
37. Xu, X., Kumakura, M., Kaku, E. & Takahashi, M. Odd mating-type substances may work as precursor molecules of even mating-type substances in *Paramecium caudatum*. *J. Eukaryot. Microbiol.* **48**, 683–689 (2001).
38. Cervantes, M. D. *et al.* Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol.* **11**, e1001518 (2013).
39. Bourc'his, D. & Voinnet, O. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**, 617–622 (2010).
40. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136**, 656–668 (2009).
41. Schoeberl, U. E. & Mochizuki, K. Keeping the soma free of transposons: programmed DNA elimination in ciliates. *J. Biol. Chem.* **286**, 37045–37052 (2011).
42. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Rev. Mol. Cell Biol.* **12**, 246–258 (2011).
43. Khurana, J. S. *et al.* Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147**, 1551–1563 (2011).
44. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Rev. Genet.* **14**, 100–112 (2013).
45. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nature Rev. Genet.* **13**, 153–162 (2012).
46. Luteijn, M. J. & Ketting, R. F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nature Rev. Genet.* **14**, 523–534 (2013).
47. Martin, A. *et al.* A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461**, 1135–1138 (2009).
48. Duret, L. *et al.* Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res.* **18**, 585–596 (2008).
49. Edgar, R., Domrachev, M. & Lash, A. E. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Malinsky, C. Ciaudo and M.-A. Félix for critical reading of the manuscript, and S. Marker and all other laboratory members for continuous support and discussions. This work was supported by the 'Investissements d'Avenir' program ANR-10-LABX-54 MEMO LIFE/ANR-11-IDEX-0001-02 Paris Sciences et Lettres* Research University and by grants ANR-08-BLAN-0233 'ParaDice' and ANR-12-BSV6-0017 'INFERNO' to E.M., L.S. and S.D., an 'Equipe FRM' grant to E.M., grants ANR-2010-BLAN-1603 'GENOMAC' and CNRS ATIP-Avenir to S.D., and National Science Foundation grant MCB-1050161 to M. Lynch (JFG). D.P.S. was supported by Ph.D. fellowships from the Erasmus Mundus program and from the Ligue Nationale Contre le Cancer. M.L.-A. was supported by Ph.D. fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche and from the Fondation de la Recherche Médicale. A.P. was supported by grant RFBR 13-04-01683a. Some strains used in this study are maintained at the Centre of Core Facilities 'Culture Collection of Microorganisms' in St Petersburg State University. The sequencing of the *mtb⁰* and *mtc⁰* MAC genomes benefited from the facilities and expertise of the high-throughput sequencing platform of IMAGiF (Centre de Recherche de Gif, <http://www.imagif.cnrs.fr>). The *mtb⁰* and *mtc⁰* transcriptomes were sequenced at the Genomic Paris Centre - IBENS platform, member of 'France Génomique' (ANR10-INBS-09-08). This study was carried out in the context of the CNRS-supported European Research Group 'Paramecium Genome Dynamics and Evolution' and the European COST Action BM1102.

Author Contributions D.P.S. did almost all of the experimental work presented here and contributed to the design of experiments. B.S. characterized mRNAs and contributed to silencing experiments and northern blot analyses. G.G. contributed to gene sequencing, plasmid construction, PCR analyses and cell line maintenance. J.-F.G. did the microarray analysis, and A.A.-F. the confocal analysis of mtA-GFP fusions. A.A., K.L. and J.-M.A. carried out the deep sequencing of small RNAs, and C.B. that of the *mtb⁰* and *mtc⁰* transcriptomes; O.A. and L.S. did the bioinformatic analyses. K.B., M.L.-A., V.T. and S.D. showed the role of scnRNA pathway genes in *mtA* promoter excision. S.B. did the *mtA* promoter dsRNA feeding experiment. A.P. contributed to the analysis of the *mtA⁰* mutant and provided *P. octaurelia* and *septaurelia* strains. M.P. contributed to the analysis of the *mtb⁰* mutant and prepared samples from the *P. octaurelia* cross, which was carried out by E.P. E.M. conceived the study and wrote the paper.

Author Information Microarray data have been deposited at the Gene Expression Omnibus database⁴⁹ under accession number GSE43436. RNA-seq data (transcriptomes of *mtb⁰* and *mtc⁰* mutants) have been deposited in the European Nucleotide Archive (EBI) under accession number ERP002291. Small RNA sequences have been deposited at the EBI under accession number ERP001812. The *mtA*, *mtb* and *mtC* sequences of all strains and species studied have been deposited at GenBank under accession codes KJ748544–KJ748569. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M. (emeyer@biologie.ens.fr).

METHODS

Paramecium strains and cultivation. Unless otherwise stated, all experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*. The mutant strains *mtA*^O, *mtB*^O and *mtC*^O, as well as strain 32 of *P. tetraurelia*, were from the stock collection of the Centre de Génétique Moléculaire in Gif-sur-Yvette, France. The *mtA*^O mutant was independently ordered from ATCC (ATCC number 30762). Strains 138 and GFg-1 of *P. octaurelia* were from the stock collection of the Institute of Systematics and Evolution of Animals, Krakow, Poland. Strain 227 from *P. septaurelia* was from the stock collection of the Centre of Core Facilities 'Cultivation of Microorganisms' in St Petersburg State University, Russia. Strain 38 of *P. septaurelia* was ordered from ATCC (ATCC number 30575). Cultivation and autogamy were carried out at 27 °C (unless otherwise stated) as described^{50,51}.

DNA and RNA extraction, Southern and northern blots. DNA and RNA samples were typically extracted from 200- to 400-ml cultures of exponentially growing cells at <1,000 cells ml⁻¹ or of autogamous cells at 2,000–4,000 cells ml⁻¹ as previously described⁵. Small-scale DNA samples were prepared from ≤1,000 cells using the NucleoSpin Tissue kit (Macherey-Nagel). The TRIzol (Invitrogen) RNA extraction procedure was adapted for small-scale cultures (~20 ml) of individual transformed clones during sexual reactivity. Electrophoresis and blotting were carried out according to standard procedures. For small-RNA northern blots, small RNAs were enriched by PEG8000 precipitation, and the equivalent of 75 µg of total RNA for each time point was run on a 15% polyacrylamide-urea gel, transferred on Hybond NX membranes, and chemically crosslinked⁵².

Microarray expression data. Expression data were obtained from single-channel NimbleGen microarrays covering all 39,642 annotated genes, with six different 50-mer probes per gene. Raw signals were processed using the standard RMA method⁵³. This includes a first step of background subtraction for each array, followed by between-array normalization which was carried out using the `normalizeBetweenArrays` function from the `limma` package⁵⁴. The latter step adjusts signals so that expression values have similar distributions in the two arrays considered in the analysis. The expression level of each gene was taken as the median signal from the six probes. The microarray platform has been described in more detail elsewhere⁵⁵.

Alignment and prediction programs. DNA and protein sequences were aligned using the MUSCLE software on the phylogeny.fr website⁵⁶. Prediction of protein localization and transmembrane protein topology used the PSORT II⁵⁷ and PolyPhobius⁵⁸ servers.

Mating-type tests. Testers were prepared from cell lines of known mating types by re-feeding ~1,000 autogamous cells in tubes with 4 ml of 0.2 X WGP medium bacterized with *Klebsiella pneumoniae* (light medium) and incubating overnight at 27 °C. The next day, tubes were re-fed with 8 ml of light medium and again incubated overnight at 27 °C. The following day, reactive cells concentrated near the top of the tube were collected (~1.5 ml per tube), checked by mixing aliquots with reactive cells of the complementary type, and used in mating type tests. Mass post-autogamous progenies to be tested were made reactive in the same way. To test individual clones, single karyonides (out of autogamy or conjugation) were isolated in 250 µl of light medium and incubated until starved. They were then re-fed with 250 µl and tested the next day.

RNAi-mediated gene silencing by dsRNA feeding and microinjections. Procedures for RNAi and DNA microinjection into the MAC, as well as inserts used to silence *scnRNA*-pathway genes and *PGM* and the procedure to silence genes during autogamy, have been described^{45,11,12,27,59,60}. To silence genes during sexual reactivity, cells were made reactive as described under 'Mating-type tests', except that the second re-feeding with 8 ml was done with dsRNA feeding medium instead of normal WGP medium.

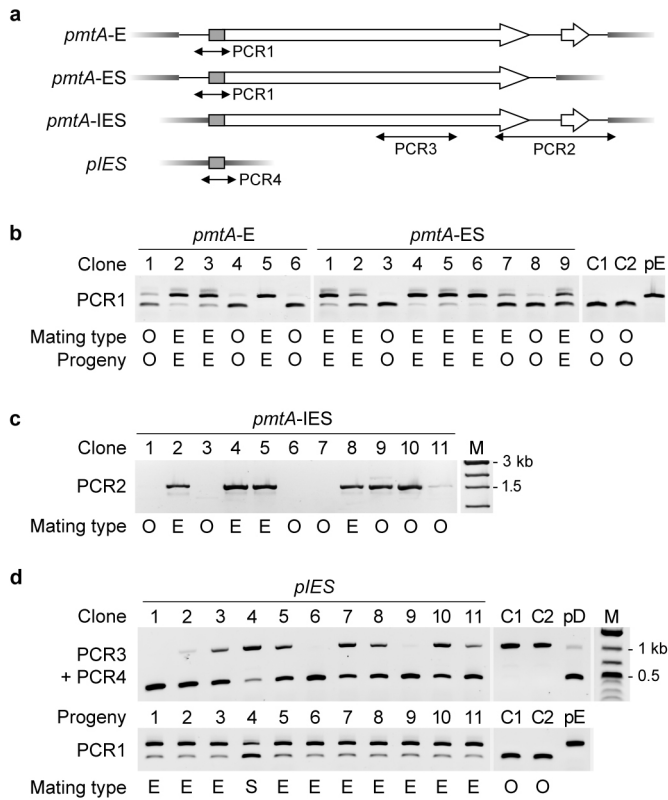
Plasmid constructs and probes. The plasmids used in this study are described in the relevant figures or Extended Data. The *mtA*-GFP fusion transgene used in Fig. 2d–f is a modified version of *pmtA*-ES (Extended Data Fig. 1) in which the EGFP coding sequence, flanked by Gly-Ser-Gly-Gly and Gly-Gly linkers, was inserted in place of *mtA* residue Asn 1002. The *pmtA*²²⁷ transgene used in Fig. 5c is the *P. septaurelia* equivalent of *pmtA*-ES, with complete flanking intergenic regions (without the downstream unannotated gene). Complete sequences of all plasmids are available on request. The *mtA* probe used to hybridize northern blots covered the whole coding sequence, except on the left panel of Fig. 1c where it spanned positions 415–1010 of the corrected gene model (PTETG5300016001, see ParametriumDB). The *P. septaurelia* probe in Fig. 5b and Extended Data Fig. 8d covered region 78–2245 of the *mtA*³⁸ gene model. For other genes, probes covered the indicated regions of each gene model: *mtAL1*, 3053–3599 of GSPATG00025159001; *mtAL2*, 3328–3790 of GSPATG00002922001; *PTTW109*, 50–439 of GSPATG00020796001; *NOWA2*, 2513–3330 of GSPATG00001668001; *PGM*, 1540–2390 of GSPATG00016627001; *PTTW110*, 4–264 of GSPATG00009468001. Oligonucleotide probes for scaffold-22 endogenous siRNAs (Extended Data Fig. 5c) have been described⁶¹.

Confocal analysis of *mtA*-GFP localization. Cells were fixed in 2% paraformaldehyde in PBS buffer and rinsed in the same buffer supplemented with 3% BSA. All subsequent steps were performed in this buffer: cells were incubated with a commercial anti-GFP rabbit serum (Molecular Probes A6455, Invitrogen) at a 1/1,000 dilution, rinsed twice, incubated with an Alexa 568-coupled anti-rabbit antibody (Molecular Probes goat anti-rabbit IgG(H+L) A11011, Invitrogen), rinsed in 1 µg ml⁻¹ Hoechst 33258, mounted in Citifluor (Citifluor Ltd London), and observed under a Leica Confocal SP8 microscope with 405, 488 and 552 laser line excitations for blue, GFP and DIC, and red detection, respectively. Images were merged using ImageJ and Adobe Photoshop.

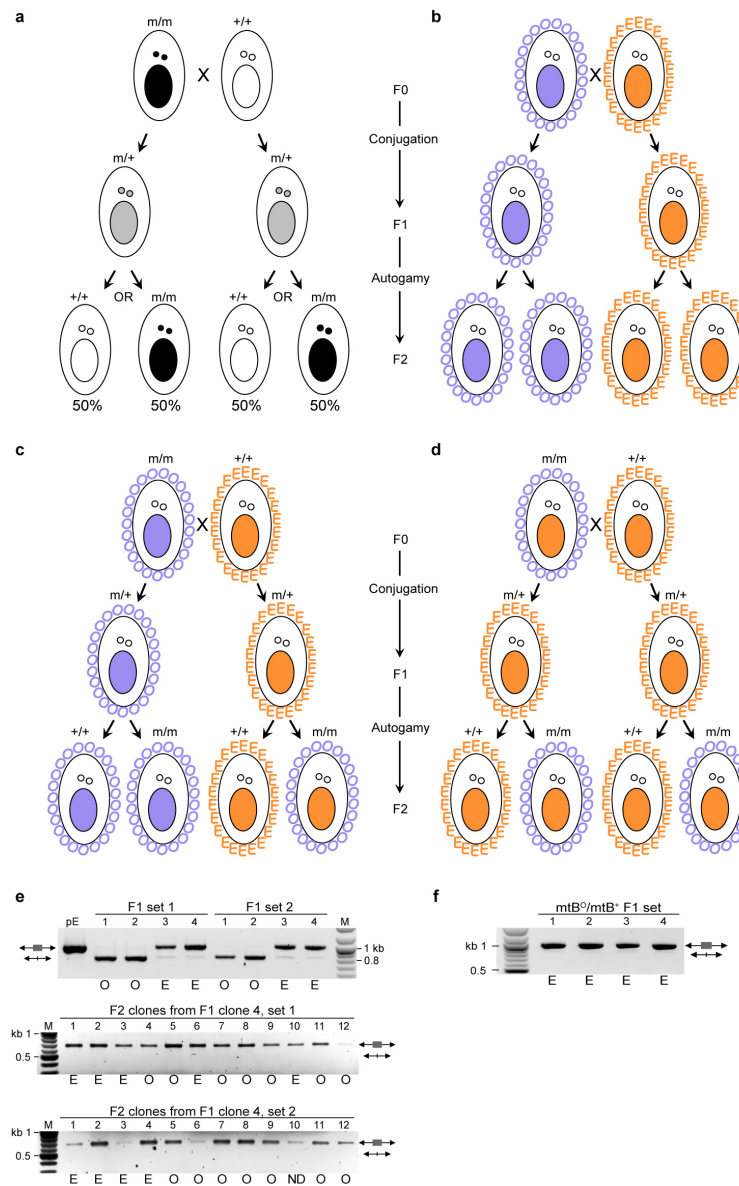
Deep sequencing of *scnRNAs*. Small-RNA libraries were previously constructed using 24-nt adaptors¹². The corresponding PCR products were extended with 6 degenerate nucleotides at each end and ligated to Illumina adaptors essentially following Illumina's recommendations (TruSeq DNA sample prep kit protocol) before sequencing on the GAIIx system (100-nt single reads) (Illumina, USA).

Whole-genome transcriptome profiling. Eight Illumina sequencing libraries were constructed (two biological replicates for *mtB*^O, two for *mtC*^O, and two wild-type controls for each) using the Illumina TruSeq RNA Sample Preparation kit. Approximately 30 million 101-nt read pairs were obtained for each library with the Illumina HiSeq 1000 sequencing system. Reads were processed to remove adapters and trimmed for quality before mapping to the strain 51 genome^{3,62} using BWA software⁶³. The fragments that mapped to each annotated gene⁶² were counted only if both reads in the pair mapped to the same gene and only if the mapping for both reads was unique in the genome. Fragments were only counted if neither read in the pair had more than two mismatches. Fragment counts per gene were obtained from the BWA output using Samtools⁶⁴ and a custom Perl script. Surface antigen genes were excluded from the analysis as the cell cultures were not controlled for serotype and surface antigen mRNA can account for 3% of total cellular mRNA. Differential gene expression was determined from fragment counts using the DESeq package⁶⁵. *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure.

50. Beisson, J. *et al.* Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb. Protoc.* **2010**, <http://dx.doi.org/10.1101/pdb.prot5362> (2010).
51. Beisson, J. *et al.* Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisoilation. *Cold Spring Harb. Protoc.* **2010**, <http://dx.doi.org/10.1101/pdb.prot5361> (2010).
52. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nature Protocols* **3**, 1077–1084 (2008).
53. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
54. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
55. Arnaiz, O. *et al.* Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genom.* **11**, 547 (2010).
56. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
57. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–35 (1999).
58. Kall, L., Krogh, A. & Sonnhammer, E. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21** (suppl. 1), 251–257 (2005).
59. Beisson, J. *et al.* DNA microinjection into the macronucleus of *Paramecium*. *Cold Spring Harb. Protoc.* **2010**, <http://dx.doi.org/10.1101/pdb.prot5364> (2010).
60. Beisson, J. *et al.* Silencing specific *Paramecium tetraurelia* genes by feeding double-stranded RNA. *Cold Spring Harb. Protoc.* **2010**, <http://dx.doi.org/10.1101/pdb.prot5363> (2010).
61. Marker, S., Le Mouél, A., Meyer, E. & Simon, M. Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.* **38**, 4092–4107 (2010).
62. Aury, J. M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178 (2006).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
64. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
66. Brygoo, Y., Sonneborn, T. M., Keller, A. M., Dippell, R. V. & Schneller, M. V. Genetic analysis of mating type differentiation in *Paramecium tetraurelia*. II. Role of the micronuclei in mating-type determination. *Genetics* **94**, 951–959 (1980).
67. Brygoo, Y. Genetic analysis of mating-type differentiation in *Paramecium tetraurelia*. *Genetics* **87**, 633–653 (1977).
68. Kapusta, A. *et al.* Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.* **7**, e1002049 (2011).



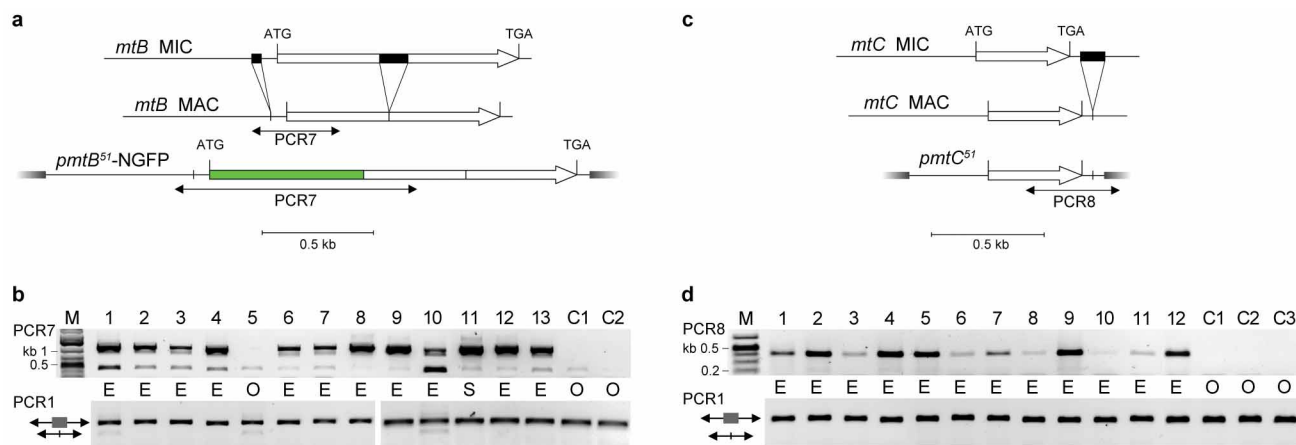
Extended Data Figure 1 | The 195-bp segment contains the *mtA* promoter and ensures maternal inheritance of its own retention. **a**, Plasmids used for microinjection into the MAC of O cells. The thick grey lines on either side of inserts represent vector sequences. Plasmid *pmtA-E* contains the E MAC version with entire intergenic regions; *pmtA-ES* lacks the downstream unannotated gene. *pmtA-IES* lacks intergenic sequences upstream of the 195-bp segment. *pIES* contains only the 195-bp segment. The PCRs used for testing transformation are shown. **b**, Transformation of O cells with *pmtA-E* or *pmtA-ES*. PCR1 amplifies the promoter-containing form from either plasmid and the promoter-excised form from the endogenous *mtA* gene; relative amounts provide an indication of plasmid copy numbers. The mating types of injected clones and of their mass post-autogamous progeny are indicated. C1 and C2, uninjected controls; pE, control PCR on *pmtA-E*. **c**, Transformation of O cells with *pmtA-IES*. The plasmid-specific PCR2 (1,533 bp) identifies transformed clones. The O type of clones 9 and 10 could be due to transgene silencing. **d**, Transformation of O cells with *pIES*. Injected clones were tested with a duplex PCR: PCR4 amplifies a 450-bp product from the plasmid and PCR3 a 1,035-bp product from the endogenous *mtA* gene, hardly detectable in high-copy transformants. C1 and C2, uninjected controls; pD, control duplex PCR on an equimolar mix of *pIES* and *pmtA-E*. PCR1 revealed that the 195-bp segment was retained in the mass progeny of transformed clones; the selfer phenotype (S) of the mass progeny from clone 4 probably reflects heterogeneity among individual post-autogamous clones.



Extended Data Figure 2 | Genetic analysis of mutations affecting mating-type expression.

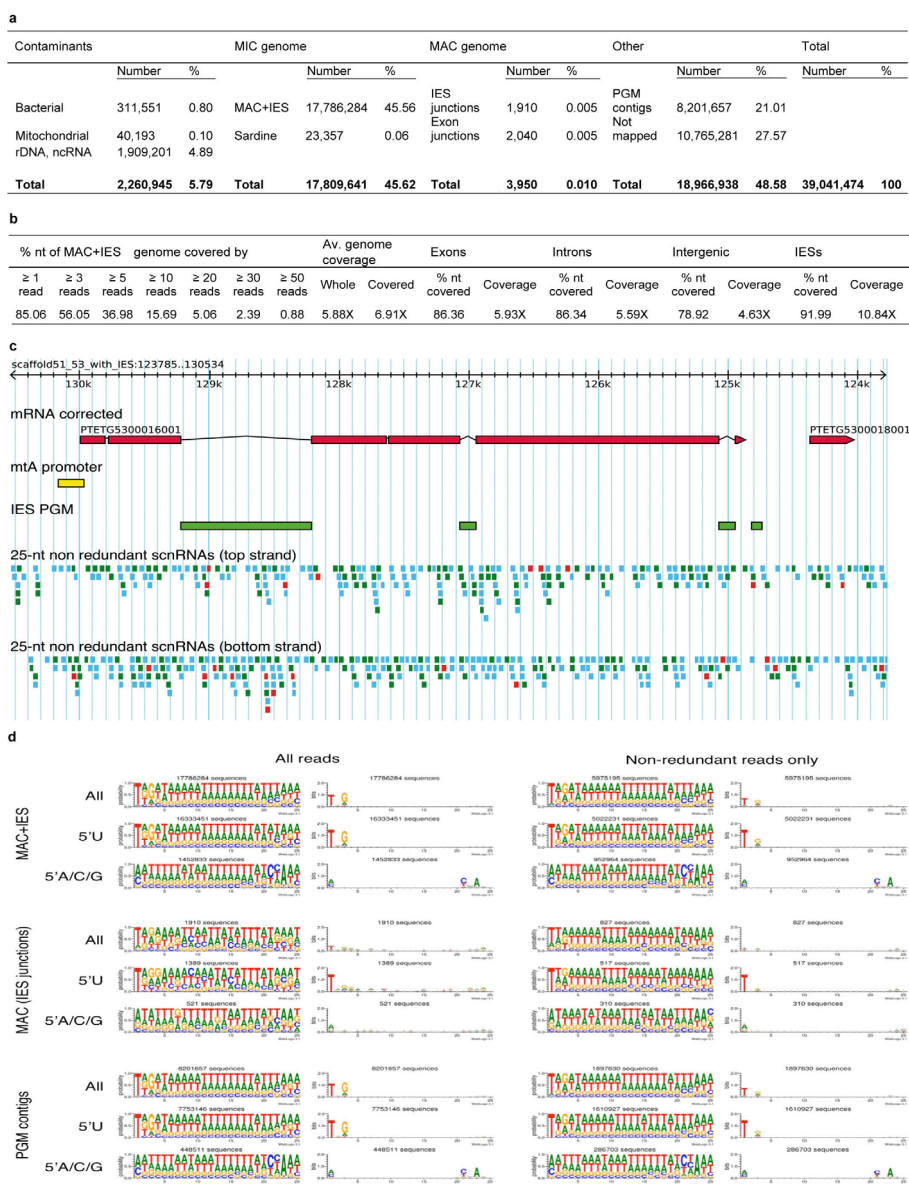
a, Mendelian segregation of a pair of alleles. In both conjugation and autogamy, the two MICs undergo meiosis, but only one haploid product is retained in each cell; an additional mitosis then produces two identical gametic nuclei. During conjugation of genetically different cells (m/m and $+/+$), reciprocal exchange of one gametic nucleus, followed by karyogamy, therefore results in genetically identical zygotic nuclei in the two exconjugants. The drawing shows the heterozygous MICs and MACs that develop from copies of the zygotic nuclei in each of the F₁ cells. During autogamy, the two identical gametic nuclei fuse together, resulting in entirely homozygous zygotes; post-autogamous F₂ progeny of heterozygotes have a 50% probability to keep each of the parental alleles. **b**, Maternal (cytoplasmic) inheritance of mating types. Coloured Os and Es around the cells indicate the mating type expressed by each vegetative clone; the MACs are coloured to symbolize their mating-type determination states (O, blue; E, orange). Because little cytoplasm is exchanged during conjugation, the parental MACs independently condition zygotic MACs for the same mating types, resulting in cytoplasmic inheritance: the F₁ derived from the O parent is determined for O in 94% of cases, and the F₁ derived from the E parent is E in 98% of cases⁶⁶. The frequency of mating-type reversal is much lower at autogamy: $<1/50,000$ in the O-to-E direction, and $1/3,000$ in the E-to-O direction⁶⁷. **c**, Cross of O-determined expression mutants mtA^O , mtB^O and mtC^O to wild-type E cells. F₂ mutant homozygotes formed in the E cytoplasmic lineage express mating type O as a result of the mutation, but remain determined for E (orange MAC). **d**, Cross of O-expressing, E-determined mutants mtA^O , mtB^O and mtC^O (as produced in c) to wild-type E

cells. E determination of the mutant parent is revealed by E determination and expression of the derived F₁, which has received the wild-type allele. In this cross all cells are determined for E, and the expressed mating type simply depends on genotype. **e**, mtA promoter retention in the cross shown in c, illustrated here for a cross of the mtA^O mutant with wild-type E cells. Top panel: PCR analysis (PCR5) of two sets of 4 mtA^O/mtA^+ F₁ karyonides (after fertilization, two new MACs develop in each cell from copies of the zygotic nucleus, and then segregate to daughter cells, called karyonides, at the first cellular division; each pair of conjugants thus gives rise to 4 F₁ karyonidal clones). Each set contained 2 O clones that had excised the mtA promoter, and probably derived from the mutant O parent, and two E clones that retained it, and probably derived from the wild-type E parent (mating types are indicated below each lane). Bottom panels: after autogamy of two mtA^O/mtA^+ F₁ heterozygotes of mating type E (clone 4 in each set), 12 F₂ homozygous progeny were isolated for each, grown and tested for mating-type expression and for mtA promoter retention using PCR6 (Supplementary Table 6), which amplifies products of 665 bp and 470 bp from the promoter-containing and promoter-excised versions, respectively. All clones retained the promoter, although a fraction of them (14 of 23) expressed mating type O, as expected for mtA^O homozygotes. ND, not determined. **f**, mtA promoter retention in F₁ heterozygotes from the cross shown in d, illustrated by a typical set of 4 mtB^O/mtB^+ F₁ karyonides from the cross of an E-determined, O-expressing mtB^O homozygote (as produced in c) to wild-type E cells. PCR5 showed that the mtA promoter was now retained in F₁ karyonides from both parents; all were of mating type E.



Extended Data Figure 3 | Complementation of the *mtB*^O and *mtC*^O phenotypes with the wild-type alleles of GSPATG00026812001 and GSPATG00009074001, respectively. **a**, Structure of the MIC and MAC versions of the *mtB* gene, and of the GFP fusion transgene used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains two IESs (black boxes). Plasmid *pmtB*⁵¹-NGFP contains the MAC version with complete intergenic regions, and the EGFP coding sequence was fused at the 5' end of the *mtB* coding sequence. Thick grey lines on either side represent plasmid vector sequences. **b**, PCR analysis and mating types of E-determined *mtB*^O mutant clones transformed with *pmtB*⁵¹-NGFP. PCR7 (top panel) amplifies products of 1,148 bp from the plasmid, and of 419 bp from the endogenous *mtB* gene (Supplementary Table 6). The relative abundance of the two products gives an indication of plasmid copy number in each clone. C1 and C2, uninjected control clones. Mating types are indicated below each lane.

Clones containing detectable plasmid amounts expressed mating type E, indicating that the GFP fusion protein is functional; the selfing phenotype (S) of clone 11 may be due to some cells having lost the plasmid. PCR1 (bottom panel) confirmed that all clones retained the *mtA* promoter. **c**, Structure of the MIC and MAC versions of the *mtC* gene, and of the plasmid used for complementation. The coding sequence (open arrow) is shown with the complete upstream and downstream intergenic regions. The MIC version contains one IES (black box). Plasmid *pmtC*⁵¹ contains the MAC version with 349 bp and 98 bp of upstream and downstream intergenic sequences, respectively. The plasmid-specific PCR8 amplifies a 419-bp product (Supplementary Table 6). **d**, PCR analysis and mating types of E-determined *mtC*^O mutant clones transformed with *pmtC*⁵¹. PCR8 (top panel) shows that all positive clones expressed mating type E. C1, C2 and C3, uninjected control clones. PCR1 (bottom panel) confirmed that all clones retained the *mtA* promoter.

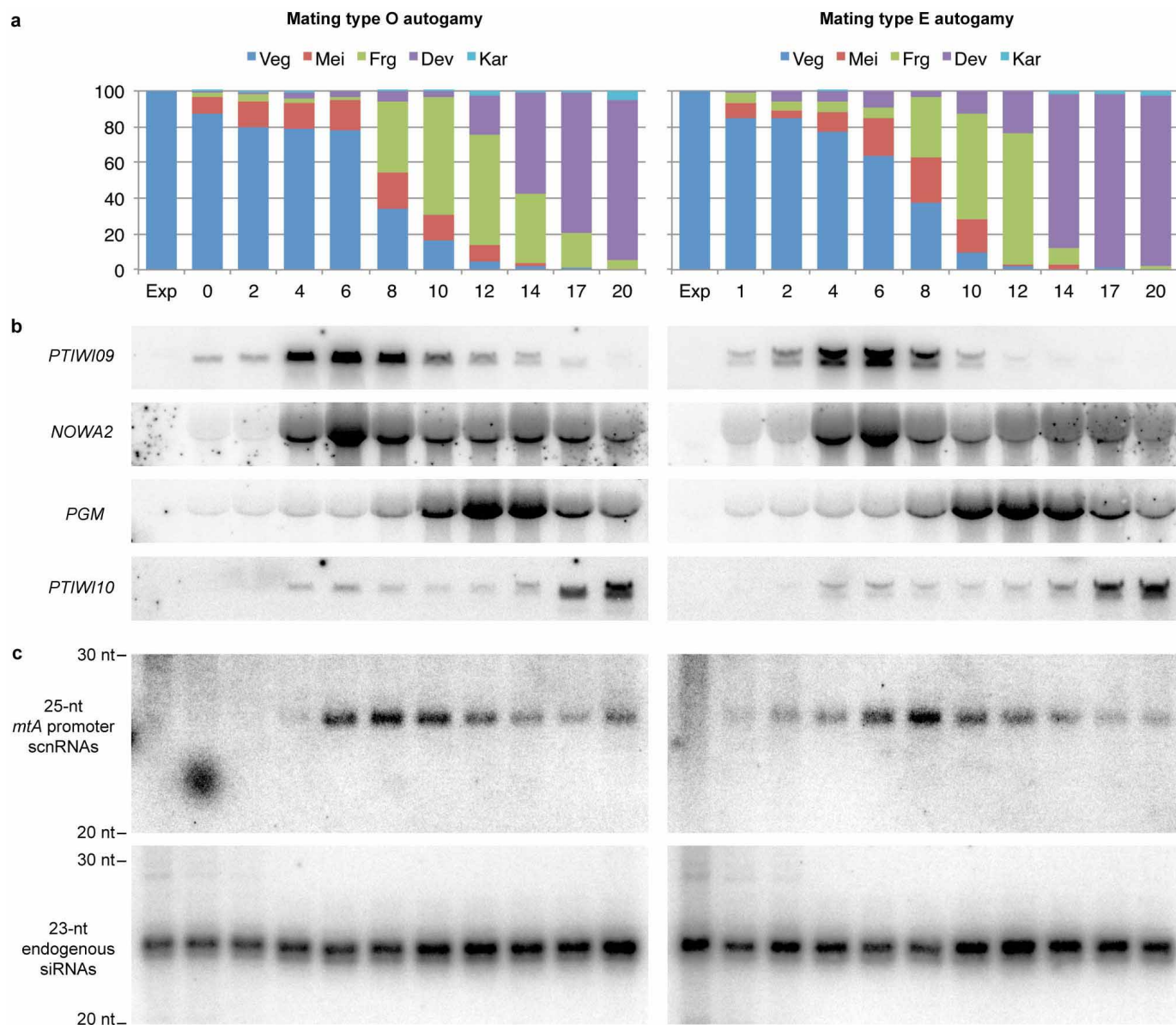


Extended Data Figure 4 | Deep sequencing of scnRNAs from an early conjugation time point (early meiosis). A total of 39,041,474 small-RNA sequences of 25 nucleotides in length were obtained by Illumina sequencing of four libraries previously constructed from gel-purified molecules migrating at 23, 24, 25, or 26 nucleotides (from the 7hND7 total RNA sample¹²).

a, 25-nucleotide reads were first mapped (no mismatch allowed) on possible contaminant sequences (genomes of bacteria commonly found in cultures, *P. tetraurelia* mitochondrial genome, *P. tetraurelia* rDNA and other non-coding RNAs). The remainder was then mapped on known MIC sequences (the 'MAC+IES' genome, and 9 individual copies of the Sardine transposon³). To determine whether any of the remaining reads could correspond to IES excision junctions or to spliced transcripts, they were then mapped on the MAC genome and on the genome-wide set of spliced transcripts. Very few hits were found and these did not show the characteristic 5'-UNG signature of scnRNAs (see d), suggesting that these molecules represent longer forms of endogenous siRNAs and/or could be mapped because of IES or intron annotation errors. Of the remaining unmapped reads (~49%), close to one-half could be mapped on 'PGM contigs', a ~25-Mb preliminary assembly of MIC-specific sequences that are not collinear to MAC chromosomes³ and are thus likely to represent bona fide scnRNAs. **b**, Statistics about the coverage of the 'MAC+IES' genome (17,786,284 reads). The average coverage is similar for exons, introns and intergenic regions, but is ~2-fold higher for IES sequences. This may mean either that scnRNAs are initially produced in higher amounts from IESs, or that active degradation of scnRNAs homologous to MAC sequences is already under way at this early stage. **c**, Mapping of 25-nucleotide reads on the MIC

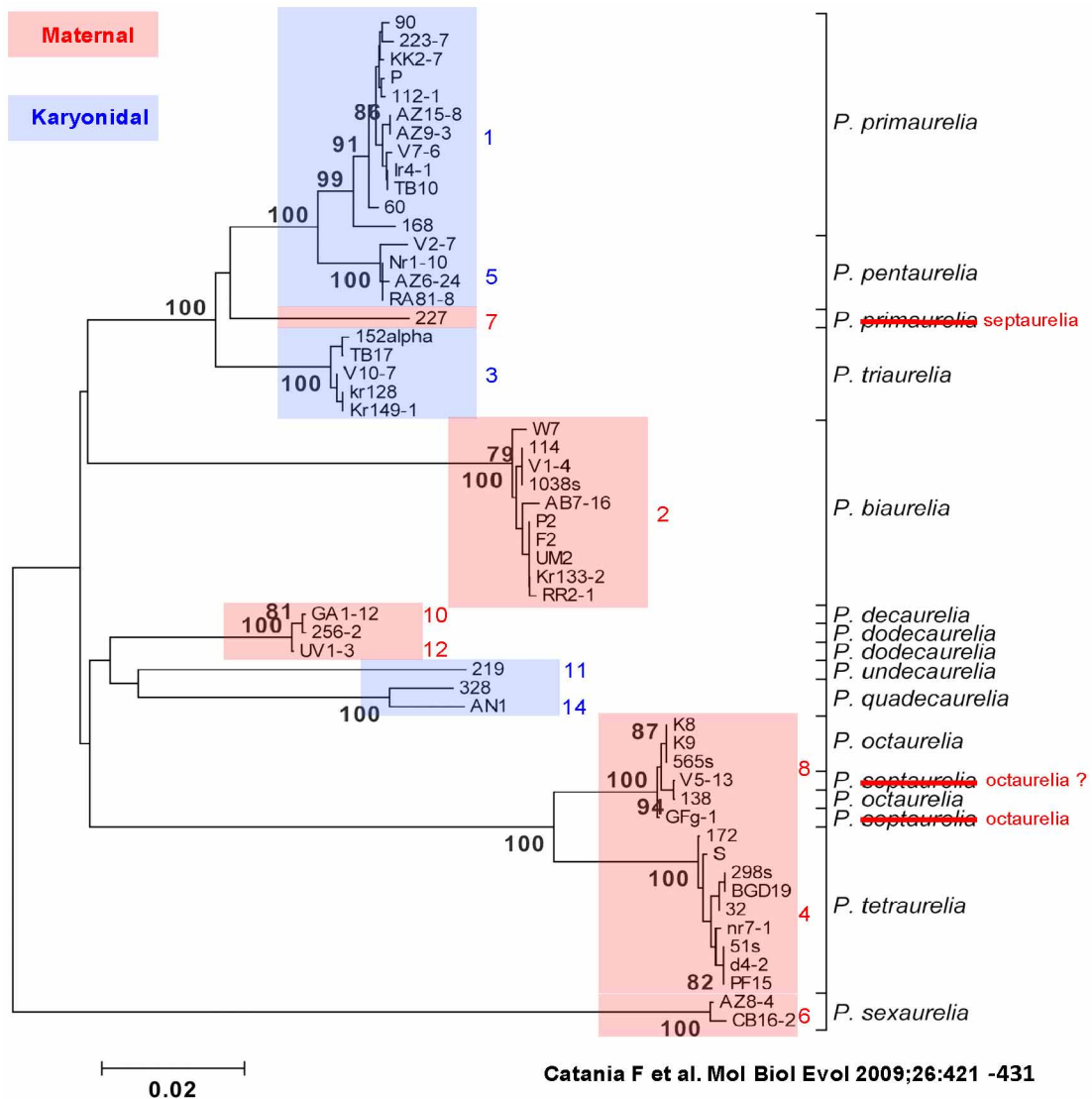
version of the *mtA* gene region. Coding sequences of the corrected *mtA* gene model (PTETG5300016001, after correction of assembly indels and re-annotation) and of the short gene downstream (PTETG5300018001) are shown as red boxes interrupted by introns and IESs. The yellow box represents the 195-bp segment of the *mtA* promoter that is excised in O cells; the 4 IESs are shown as green boxes. 25-nucleotide reads mapping on the top or bottom strands of the region are colour-coded to indicate the number of times each one was sequenced: blue, one read; green, 2–9 reads; red, ≥ 10 reads.

d, Compositional profiles (nucleotide frequency on the left, and deviation from randomness on the right) of reads mapping to the MAC+IES genome, to the MAC genome, or to 'PGM contigs'. For each set, logos are shown for all reads (left), or for the non-redundant subset only (right). For the 'MAC+IES' and 'PGM contigs' sets, the logos computed from all reads clearly show the 5'-UNG signature typical of scnRNA guide strands, as is the case for the major subset of reads starting with U (5'U), while the minor subset of sequences not starting with U (5'A/C/G), which may represent the steady-state amount of passenger strands, shows the complementary signature CNA at positions 21–23. Deviation from randomness is greater when computed from all reads than when computed from the non-redundant subset only, indicating that molecules with the signature are intrinsically produced in higher amounts, or are more stable. The small set of reads that mapped only to the MAC genome (putative IES excision junctions) does not show a clear 5'-UNG signature, suggesting that most of those are not scnRNAs. (The same is true of the small number of reads mapping to exon–exon junctions (not shown).)



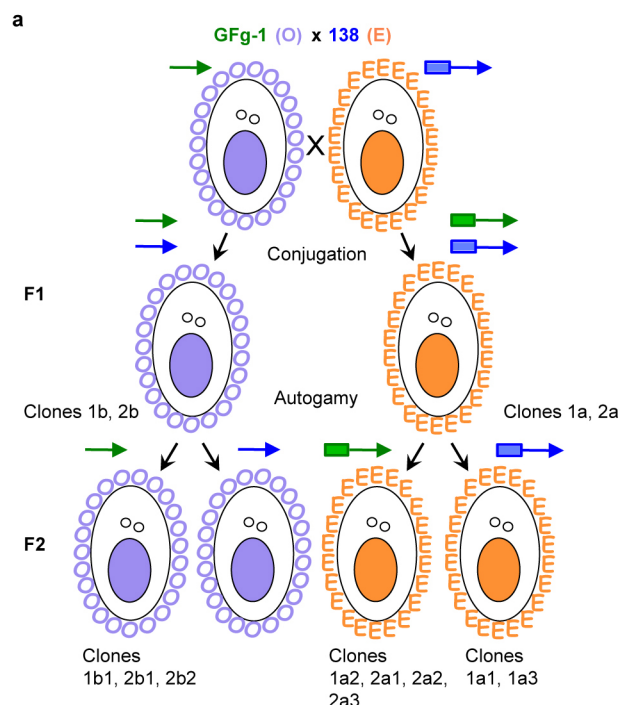
Extended Data Figure 5 | Northern blot analysis of *mtA*-promoter scnRNAs during autogamy of O or E cells. Mass cultures were allowed to starve, and RNA samples were extracted during exponential growth (Exp) and then at different times (0–20 h) after the appearance of the first meiotic cells. Cells become committed for autogamy at a fixed point of the cell cycle, so that the best synchrony that can theoretically be achieved is the duration of one cell cycle; in these experiments, the time between the first and the last cells to begin meiosis was ~12 h. **a**, Proportions of cells in different cytological stages at each time point, as determined by DAPI staining. Veg, vegetative cells; Mei, meiosis (crescent stage, meiosis I, meiosis II); Frg, cells with fragmented old MAC but new MACs not yet clearly visible; Dev, cells with two clearly visible developing new MACs; Kar, cells after the karyonidal division, with only one developing new MAC. **b**, Northern blot analysis of mRNAs for early (*PTIWI9*, *NOWA2*), middle (*PGM*), or late (*PTIWI10*) genes. The same blots were hybridized successively with the 4 probes. **c**, Northern blot analysis of small

RNAs. The top panels show hybridization with the 195-bp *mtA*-promoter probe, revealing accumulation of 25-nucleotide scnRNAs slightly later than expression of the meiosis-specific genes *PTIWI9* and *NOWA2*. As a control, the same blots were rehybridized with an oligonucleotide probe specific for a cluster of 23-nucleotide endogenous siRNAs on scaffold 22^{12,61}, which are abundantly produced at all stages of the life cycle (bottom panels). Quantification of the *mtA*-promoter scnRNA signal and normalization with the siRNA signal did not reveal any significant difference in their amount or timing between the two mating types (not shown). Previous studies showed that the double-strand breaks that initiate IES excision in the new MACs start being detectable before the maximum of expression of the putative endonuclease *PGM*^{5,68} (no later than 10 h in these time courses). A PCR analysis of post-autogamous DNA samples confirmed that the *mtA* promoter was fully excised in mating type O, and fully maintained in mating type E (not shown).



Extended Data Figure 6 | Phylogenetic tree of different strains from most *P. aurelia* species based on sequence polymorphisms in three nuclear genes. This figure is modified with permission from figure 2 of Catania *et al.* Genetic diversity in the *Paramecium aurelia* species complex *Mol. Biol. Evol.* 2009, 26, 421–431 (ref. 32). Wrong species assignment of some strains studied

here are corrected in red. Strain V5-13 was probably mis-assigned to *P. septaurelia* through the same error as for strain GFG-1 (see Extended Data Fig. 7). Species showing maternal inheritance or random determination of mating types are highlighted in red and blue, respectively.

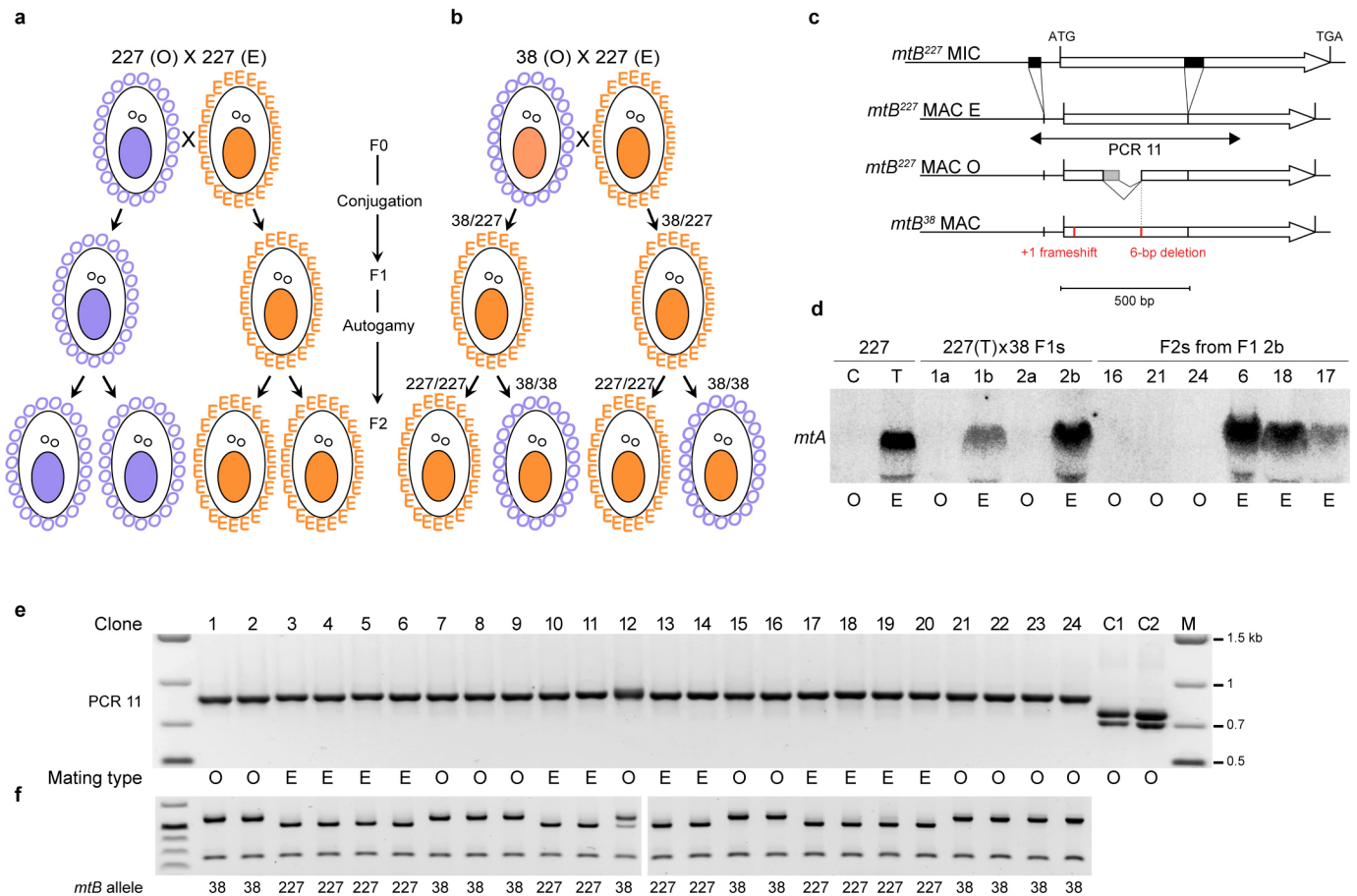


b

	Mitochondrial COXII variant	mtA genotype	mtA promoter status	Inferred mating type
F1 pairs				
1a	138	ND	ND	E
1b	GFg-1	138/GFg-1	excised	O
2a	138	ND	ND	E
2b	GFg-1	138/GFg-1	excised	O
F2 clones, E side				
1a1	138	138/138	retained	E
1a2	138	GFg-1/GFg-1	retained	E
1a3	138	138/138	retained	E
2a1	138	GFg-1/GFg-1	retained	E
2a2	138	GFg-1/GFg-1	retained	E
2a3	138	GFg-1/GFg-1	retained	E
F2 clones, O side				
1b1	GFg-1	GFg-1/GFg-1	excised	O
2b1	GFg-1	GFg-1/GFg-1	excised	O
2b2	GFg-1	GFg-1/GFg-1	excised	O

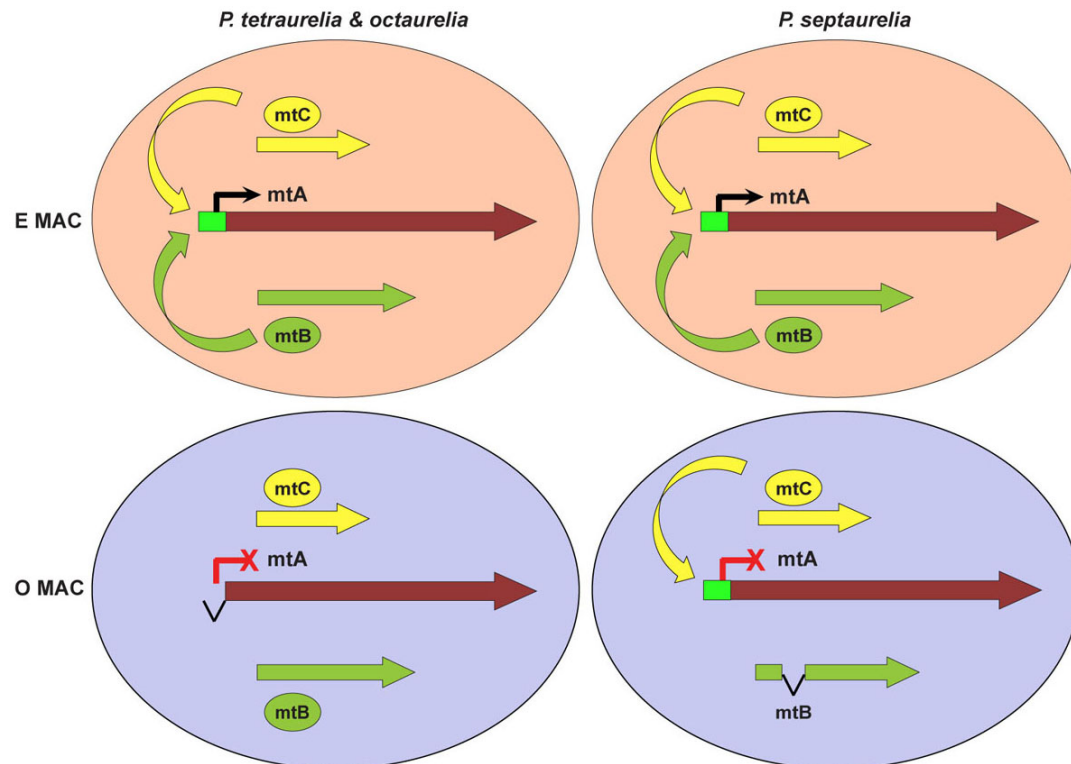
Extended Data Figure 7 | Strain GFg-1 belongs to the same species as strain 138, that is, *P. octaurelia*. GFg-1 is among a set of strains that were originally assigned to *P. septaurelia* on the basis of conjugation tests with strain 38, a reference strain for that species. However, the stock of strain 38 used in these tests was not 38, but instead some *P. octaurelia* strain, as shown by the comparison of *mtA* sequences with those from the original strain 38 obtained from ATCC (ATCC number 30575) and those from strain 138, a reference strain for *P. octaurelia*. **a**, Scheme of the cross GFg-1 (O) × 138 (E). The mating types of parents were determined by cross-agglutination with *P. tetraurelia* tester lines. The green and blue arrows beside each cell represent the *mtA* gene, colour-coded to indicate the GFg-1 and 138 alleles; the box at the 5' end symbolizes retention of the *mtA* promoter in the MAC genome on the E side of the cross. **b**, Molecular characterization of two pairs of F₁ heterozygotes, and of

some F₂ clones obtained by autogamy of these F₁ heterozygotes. The parental origin of each F₁ clone was ascertained by a sequence polymorphism in the mitochondrial *COXII* gene (PCR9, Supplementary Table 6). PCR amplification and sequencing of a segment of the *mtA* gene encompassing the promoter (PCR10, Supplementary Table 6) showed that the two F₁ clones deriving from the GFg-1 parent were heterozygotes, and that the *mtA* promoter was precisely excised from both alleles (see Supplementary Data 1a). ND, not determined. Analysis of 6 viable F₂ clones obtained by autogamy of the F₁ clones deriving from the 138 parent showed that 4 of them were homozygous for the GFg-1 allele, whereas the other two were homozygous for the 138 allele; the *mtA* promoter was retained in all cases. The evidence for successful genetic exchange between strains GFg-1 and 138 and for viable recombinant F₂ progeny demonstrates that these strains belong to the same species; that is, *P. octaurelia*.



Extended Data Figure 8 | Genetic and molecular analysis of the cross between strains 227 and 38 of *P. septaurelia*. **a**, Maternal inheritance of mating types in the wild-type strain 227. **b**, Strain 38 is genetically restricted to O expression, but constitutively determined for E. The mutation identified in strain 38 is of particular interest because it affects both the expression and the inheritance of mating types, which suggests that it lies in a gene that controls mating-type determination through an alternative rearrangement. Indeed, known *P. tetraurelia* mutations fall in two distinct categories. mtA^O , mtB^O and mtC^O prevent expression of type E but have no effect on the rearrangement that determines mating types or on its maternal inheritance, whereas mtE^E lies in a *trans*-acting factor required for a subset of rearrangements during MAC development but has no effect on the expression of mating types during sexual reactivity. The only type of mutation that can be envisioned to affect both expression and determination/inheritance would be a mutation preventing expression of a functional protein required for E expression, and at the same time preventing in *cis* (by destroying a potential Pgm cleavage site) the rearrangement that normally inactivates this gene in the MAC of wild-type O cells. The mt^{XIII} allele of strain 38 restricts cells to O expression in a recessive manner, but also has a maternal effect that enforces constitutive E determination in sexual progeny. Notably, elegant experiments showed the latter effect to be dominant²⁰; the sexual progeny of a cell carrying at least one mt^{XIII} allele can never be determined for O or transmit O determination. **c**, Sequencing of the mtB^{38} allele revealed features that may account for both effects. A frameshift mutation makes it a pseudogene, explaining the genetic restriction to O expression. In addition, a 6-bp deletion removes one of the IES-like boundaries used in the mtB^{227} allele for coding-sequence deletions in O

clones. Given the requirements of IES excision in *P. tetraurelia* and in sibling species³⁶, this can be predicted to make the same deletions impossible in the *mtB*³⁸ allele, which would explain the constitutive E determination effect. The full-length *mtB*³⁸ pseudogene in the MAC of 38 cells would indeed be expected, after a cross to 227, to protect the highly similar zygotic *mtB*²²⁷ allele against coding-sequence deletions in the derived F₁, through titration of homologous scnRNAs. **d**, Molecular analysis of the 38 × 227 cross. To verify this maternal effect, we crossed an E-expressing 227 clone (*pmtB*⁵¹-transformed clone T, same as in Fig. 5b; C, uninjected control) with strain 38, and F₁ heterozygotes were tested for mating types and for *mtA* expression by northern blotting. As expected, F₁ heterozygotes deriving from the 38 parent (1b and 2b, as determined by sequencing of a mitochondrial polymorphism) were E and expressed *mtA*, indicating that the incoming *mtB*²²⁷ allele had been rearranged into a functional, full-length form in the MAC. After autogamy of F₁ clone 2b, 24 independent F₂ homozygotes were isolated and tested for mating types. Consistent with the Mendelian segregation of *mtB* alleles, 12 were O and 12 were E (Supplementary Table 5); northern blot analysis of 3 clones of each type showed that only E clones expressed *mtA*. **e**, All F₂ clones maintained the full-length *mtB* gene in the MAC. PCR11 (Supplementary Table 6) amplifies an 888-bp fragment from the MAC version of *mtB*³⁸, and an 893-bp fragment from the full-length MAC version of *mtB*²²⁷. C1 and C2, control PCR11 on two O clones of strain 227, showing the 806-bp and 744-bp fragments resulting from the two alternative coding-sequence deletions. **f**, Mating types co-segregate with *mtB* alleles among F₂ homozygotes. Digestion of the PCR products with AluI distinguishes the 38 and 227 alleles. *mtA* and *mtC* alleles segregated independently (Supplementary Table 5).



Extended Data Figure 10 | A general model for mating-type determination in *P. aurelia* species. In the three species examined, mating type E depends on expression of the *mtA* protein during sexual reactivity. *mtA* transcription in turn requires the *mtB* and *mtC* gene products (the requirement for *mtC* in *P. septaurelia*, and for both genes in *P. octaurelia*, remains to be verified). In

P. tetraurelia and *P. octaurelia*, mating type O is determined during MAC development by excision of the *mtA* promoter as an IES, preventing expression of the gene. In *P. septaurelia*, mating type O is determined by the excision of segments of the *mtB* coding sequence as IESs, which similarly prevents *mtA* expression.

Amygdala interneuron subtypes control fear learning through disinhibition

Steffen B. E. Wolff^{1,2*}, Jan Gründemann^{1*}, Philip Tovote¹, Sabine Krabbe¹, Gilad A. Jacobson¹, Christian Müller¹, Cyril Herry³, Ingrid Ehrlich⁴, Rainer W. Friedrich¹, Johannes J. Letzkus^{1†*} & Andreas Lüthi^{1*}

Learning is mediated by experience-dependent plasticity in neuronal circuits. Activity in neuronal circuits is tightly regulated by different subtypes of inhibitory interneurons, yet their role in learning is poorly understood. Using a combination of *in vivo* single-unit recordings and optogenetic manipulations, we show that in the mouse basolateral amygdala, interneurons expressing parvalbumin (PV) and somatostatin (SOM) bidirectionally control the acquisition of fear conditioning—a simple form of associative learning—through two distinct disinhibitory mechanisms. During an auditory cue, PV⁺ interneurons are excited and indirectly disinhibit the dendrites of basolateral amygdala principal neurons via SOM⁺ interneurons, thereby enhancing auditory responses and promoting cue–shock associations. During an aversive footshock, however, both PV⁺ and SOM⁺ interneurons are inhibited, which boosts postsynaptic footshock responses and gates learning. These results demonstrate that associative learning is dynamically regulated by the stimulus-specific activation of distinct disinhibitory microcircuits through precise interactions between different subtypes of local interneurons.

Fear conditioning is a powerful model system for investigating plasticity of neuronal circuits and the mechanisms of associative learning, as the basic underlying circuitry has been thoroughly studied over the last decades^{1–4}. The basolateral amygdala (BLA) has been identified as a key brain area involved in auditory fear conditioning, where associative synaptic plasticity at glutamatergic sensory afferents is induced by the pairing of a neutral tone (conditioned stimulus (CS)) and a mild aversive footshock (unconditioned stimulus (US)) during conditioning^{1,4,5}. As a cortex-like structure⁶, the BLA contains a variety of interneurons with marker expression, basic properties and connectivity very similar to neocortical circuits^{5,7–13}, and inhibition has been shown to have a role in fear conditioning⁸. Functionally, salient sensory stimuli lead to cell-type-specific responses in different interneuron subtypes in anaesthetized animals¹⁴. Moreover, fear conditioning and extinction have been shown to induce opposing changes in the strength of inhibitory transmission and in the expression of inhibition-related genes^{8,15–17}. However, besides the general notion that inhibition is involved in fear learning, knowledge about the role of individual interneuron subtypes is still lacking.

One of the major interneuron subclasses in cortical and cortex-like structures such as the BLA is defined by the expression of the calcium-binding protein parvalbumin (PV; also called PVALB)^{9,10,18–20}. Most PV⁺ interneurons feature a fast-spiking phenotype with narrow spike widths and high firing rates^{21,22}. PV⁺ interneurons preferentially form synapses at the perisomatic region of their target cells, thereby controlling neuronal activity and spike output^{10,18,23,24}. In addition to glutamatergic principal neurons, PV⁺ cells contact both other PV⁺ interneurons and different interneuron types^{21,25}. In the hippocampus, PV⁺ cells have been shown to interact with interneurons expressing somatostatin (SOM; also called SST) to regulate pyramidal cell output²⁶. In contrast to PV⁺ cells, SOM⁺ interneurons preferentially contact the distal dendrites²⁷, which allows them to efficiently control the impact of inputs to their target cells^{28,29}.

Here, we address the role of interneuron subtypes using a combination of optogenetic manipulations and single-unit recordings from identified BLA interneurons during fear learning. Our results reveal that BLA PV⁺ and SOM⁺ interneurons exert bidirectional control over fear acquisition resulting in differential changes in inhibition along the somatodendritic axis of principal neurons. We identify inhibition of SOM⁺ interneurons by PV⁺ interneurons as a crucial mechanism for dendritic disinhibition of principal neurons. Our findings indicate that exquisitely organized inhibitory and disinhibitory microcircuits comprising defined subtypes of interneurons have important and distinct roles during associative learning.

PV⁺ interneurons in the BLA control fear learning

To examine a possible role of PV⁺ cells in fear learning, we selectively expressed channelrhodopsin-2 (ChR2 (ref. 30)) in PV⁺ interneurons in the BLA³¹ (Fig. 1a–c; see Methods), and used bilaterally implanted optic fibres for light stimulation (Fig. 1a and Extended Data Fig. 1a). Importantly, stimulation of PV⁺ interneurons alone did not have any effect on freezing or locomotion (Extended Data Fig. 2a). Mice were conditioned to two distinct CS (tones of different frequencies, counterbalanced between experiments). Whereas PV⁺ cell activity was manipulated during one entire CS–US pairing, the other CS–US pair served as a within-animal control (Fig. 1d, left). A fear memory retrieval test on the next day without optogenetic stimulation revealed that PV⁺ interneuron activation during acquisition significantly attenuated the CS-induced freezing response when compared with the internal control CS (Fig. 1d, right and Extended Data Fig. 2b). These results indicate that the level of PV⁺ interneuron activity in the BLA can control the acquisition of conditioned fear responses.

US-induced inhibition of PV⁺ interneurons gates learning

Next, we asked whether optogenetic activation of PV⁺ interneurons specifically during the US would be sufficient to interfere with fear

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. ²University of Basel, 4000 Basel, Switzerland. ³INSERM U862, Neurocentre Magendie, 146 rue Leo Saignat, 33077 Bordeaux, France. ⁴Hertie Institute for Clinical Brain Research, 72076 Tübingen, Germany. [†]Present address: Max-Planck Institute for Brain Research, 60438 Frankfurt, Germany.

*These authors contributed equally to this work.

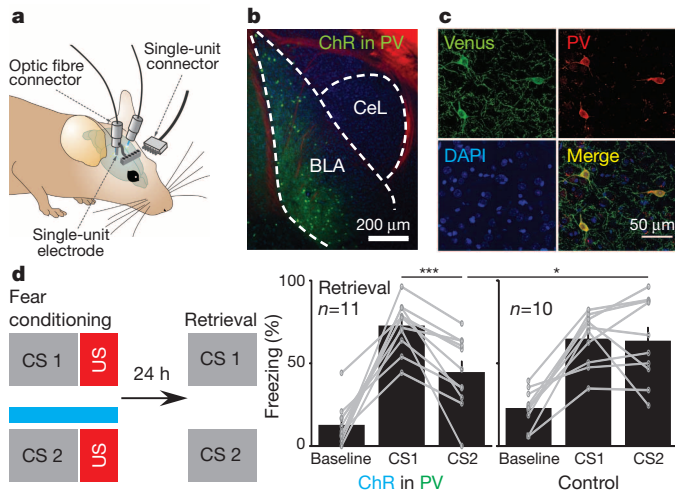


Figure 1 | BLA PV⁺ interneurons control fear learning. **a**, Optogenetic manipulation and simultaneous single-unit recordings using an optrode in freely behaving mice. **b**, Co-expression of ChR2 and Venus (green) in PV⁺ interneurons in the BLA, using PV-Cre mice and conditional AAVs. Counterstains: DAPI (blue) and NeuN (red). CeL, central lateral amygdala. **c**, Restriction of ChR2 and Venus expression (green) to PV⁺ interneurons is verified by PV immunohistochemistry (IHC) (red). **d**, Left, fear conditioning paradigm to test for effects of PV⁺ manipulation on fear learning. Animals were conditioned to two distinct auditory CSs. CS2 was paired with blue light for ChR2 activation during the entire CS-US pairing. During retrieval on the next day, both CSs were presented without light. CS order and light-pairing were counterbalanced. Right, animals displayed significantly less freezing to the light-paired CS2 during retrieval compared to CS1, indicating that PV⁺ activation during the entire CS-US pairing impairs fear acquisition. Light stimulation had no effect on control animals expressing only GFP. Values are mean \pm s.e.m. * P < 0.05, *** P < 0.001. Statistical analysis in Methods.

conditioning (Fig. 2a and Extended Data Fig. 2b). When we tested fear memory retrieval the next day without optogenetic stimulation, we observed that learning was impaired to a similar extent as observed with PV⁺ interneuron activation during the entire CS-US pairing (Fig. 1d and Extended Data Fig. 2b). This demonstrates that PV⁺ interneuron activity during the US is a crucial factor regulating fear learning. To address whether endogenous activity of PV⁺ cells is necessary for fear learning, we used conditional expression of the inhibitory opsin ARCH³² (Fig. 2a). In contrast to ChR2 manipulations, ARCH-mediated inhibition of PV⁺ interneurons during the US resulted in increased freezing levels during fear memory retrieval (Fig. 2a and Extended Data Fig. 2b). Together, these results demonstrate that fear learning is under bidirectional control of BLA PV⁺ interneurons during the US.

To relate the behavioural changes induced by these manipulations to the physiological role of PV⁺ interneurons, we used custom-built optrodes (Fig. 1a and Extended Data Fig. 1; see Methods) to perform chronic single-unit recordings from optogenetically identified PV⁺ cells³³. Units exhibiting light-induced short latency activity changes were identified as PV⁺ interneurons (Fig. 2b, c and Extended Data Fig. 3a–e; see Methods for identification criteria). In agreement with previous recordings in slices^{22,34}, optogenetically identified PV⁺ interneurons showed faster spike kinetics and higher spontaneous activity than simultaneously recorded non-light-responsive units (Fig. 2d and Extended Data Fig. 4). Recordings from identified light-responsive units ($n = 31$) revealed that firing of the PV⁺ interneuron population was reduced for the entire duration of the footshock (Fig. 2e), although a minority displayed excitation (7 out of 31, Extended Data Fig. 5a). These data are consistent with the observed behavioural effects: optogenetic activation of PV⁺ interneurons counteracted the endogenous US-induced inhibition and decreased learning, whereas optogenetic inhibition enhanced the endogenous PV⁺ inhibition resulting in enhanced fear learning (Fig. 2a and Extended Data Fig. 2b).

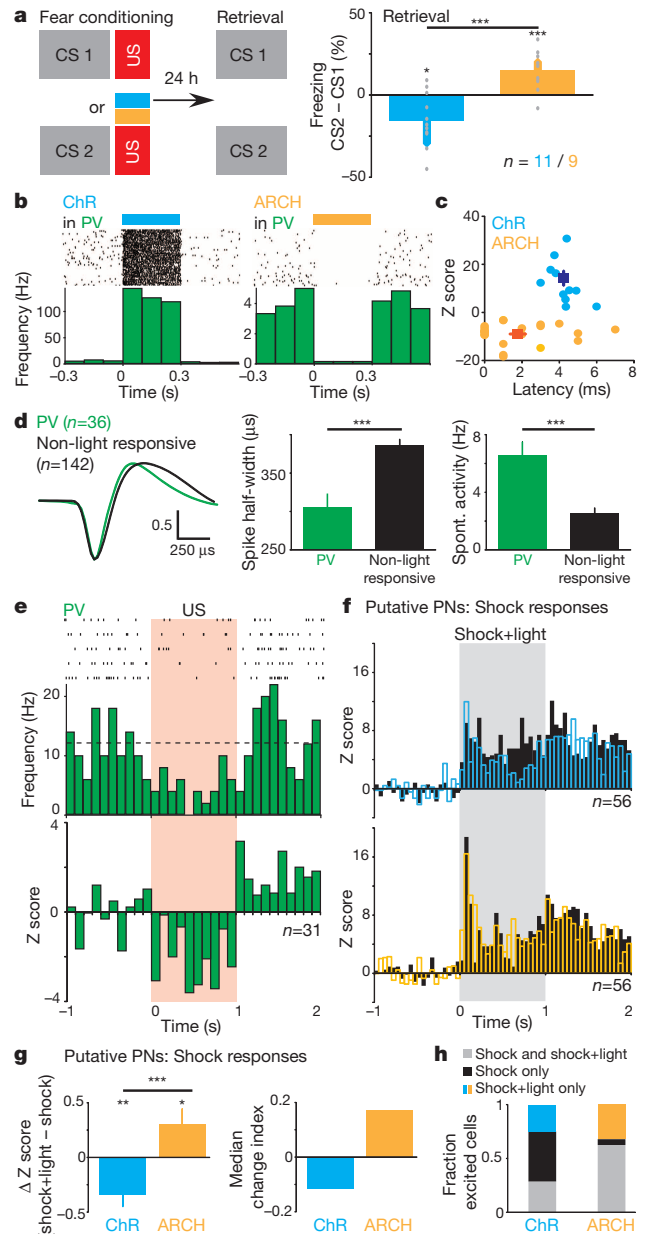


Figure 2 | Footshock-induced PV⁺ cell inhibition gates fear learning.

a, PV⁺ activity during the US regulates fear learning. Left, conditioning as in Fig. 1d, but illumination with blue or yellow light was restricted to the US. Right, PV⁺ activation and inhibition during the US had opposite effects on fear learning. Shown are differences in freezing between the two CSs. **b**, Optogenetic identification of PV⁺ interneurons in the BLA of freely behaving mice, based on light-induced changes in single-unit activity. **c**, Light-responsive units were defined as those with significant, short-latency light-induced changes in activity. **d**, Optogenetically identified PV⁺ units differ in spike width and spontaneous firing rate from simultaneously recorded non-light-responsive units, mainly representing principal neurons. **e**, Footshocks inhibit PV⁺ cell activity during conditioning (top, example response; bottom, Z-scored population response). **f**, Z-scored population responses for footshock-excited putative principal neurons (PNs) with (blue or yellow) and without (black) optogenetic stimulation of PV⁺ interneurons (for selection criteria see Methods). **g**, Left, Z-score differences in the footshock response induced by optogenetic manipulation of PV⁺ interneurons in footshock-excited putative principal neurons. Right, median change index for the footshock responses of putative principal neurons upon PV⁺ cell manipulations (see Methods). **h**, Fractions of excited putative principal neurons, which are excited either exclusively during the footshock (black), exclusively during the footshock + light (blue or yellow) or in both conditions (grey). Values are mean \pm s.e.m. unless stated otherwise. * P < 0.05, ** P < 0.01, *** P < 0.001. Statistical analysis in Methods.

Because PV⁺ interneurons mediate strong perisomatic inhibition^{22,23,34}, we speculated that US-induced inhibition of PV⁺ cells causes disinhibition of principal neurons. We therefore analysed the impact of PV⁺ interneuron activity on excitatory footshock responses in putative principal neurons (non-light-responsive units with spontaneous firing rates <3 Hz; $n = 112$)^{5,7,12,35,36}. We observed a significant decrease in excitatory footshock responses upon optogenetic activation of PV⁺ interneurons (Fig. 2f, g and Extended Data Fig. 6). In contrast, inhibition of PV⁺ interneurons significantly increased the principal neuron responses (Fig. 2f, g and Extended Data Fig. 6). In addition to affecting the strength of the footshock-induced responses, optogenetic manipulations of PV⁺ interneuron activity differentially changed the number of footshock-excited cells. Whereas PV⁺ activation led to a loss of excitation in almost half of the footshock-excited cells, PV⁺ inhibition uncovered excitatory responses in a large fraction of putative principal neurons (Fig. 2h and Extended Data Fig. 6). To address directly whether changes in the footshock-induced activity of principal neurons have an effect on learning, we expressed ARCH under the control of the CaMKII promoter and optogenetically inhibited principal neurons during the US. This manipulation significantly impaired fear learning, demonstrating the necessity of footshock-induced principal neuron activation (Extended Data Fig. 6e). Together, these results indicate that footshock-induced inhibition of PV⁺ cells regulates activity in the BLA network in a manner consistent with our behavioural observations: PV⁺ interneuron inhibition causes disinhibition and enhances US-induced activity in principal neurons, probably acting as a permissive gate for plasticity induction and fear learning.

Differential roles for PV⁺ and SOM⁺ interneurons during CS

We next addressed the question whether inhibition of PV⁺ interneurons and the consequent disinhibition of principal neurons also occur during the auditory CS which precedes the footshock US. Unexpectedly, when optogenetically manipulating the activity of PV⁺ interneurons exclusively during the CS, we found the opposite effects on fear learning compared to US-specific manipulations (Fig. 3a and Extended Data Fig. 2b). This suggests fundamentally different roles of PV⁺ interneurons in processing of CS and US.

In line with this, the PV⁺ interneuron population exhibited CS-induced excitation (Fig. 3b), although a subset displayed inhibitory responses (14 out of 36, Extended Data Fig. 5c). If PV⁺ interneurons influenced principal neurons mostly via monosynaptic perisomatic inhibition, we would expect optogenetic manipulations of PV⁺ interneurons to have an effect on learning opposite to the one that we observed.

A possible explanation for the paradoxical effect of PV⁺ interneurons on fear conditioning could be that PV⁺ interneurons interact not only with principal neurons, but in addition with other interneuron subtypes to cause disinaptic disinhibition of principal neurons. Because sensory afferents from auditory thalamus and cortex make synapses onto the dendrites of BLA principal neurons⁵, we proposed that dendrite-targeting SOM⁺ interneurons might mediate such an effect. To address this, we performed optogenetic manipulations and single-unit recordings from identified SOM⁺ interneurons using the same strategy as for PV⁺ interneurons (Extended Data Figs 1, 3, 4)³⁷. In agreement with the hypothesis that SOM⁺ interneurons could be inhibited by PV⁺ cells during the CS, we found that the behavioural consequences of optogenetic manipulations of SOM⁺ cells during the CS were opposite to those induced by PV⁺ manipulations (Fig. 3c and Extended Data Fig. 2b), and that CS presentation induced inhibition of identified SOM⁺ interneurons (Fig. 3d and Extended Data Fig. 5d). Taken together, this indicates that the CS causes subcellular reallocation of inhibition in principal neurons, with disinhibition of sensory dendritic inputs, which overrides a concomitant increase in perisomatic inhibition (see below).

Dendritic disinhibition by SOM⁺ interneurons

To address whether PV⁺ interneurons indeed make synaptic contacts onto SOM⁺ interneurons as has been demonstrated in the hippocampus²⁶

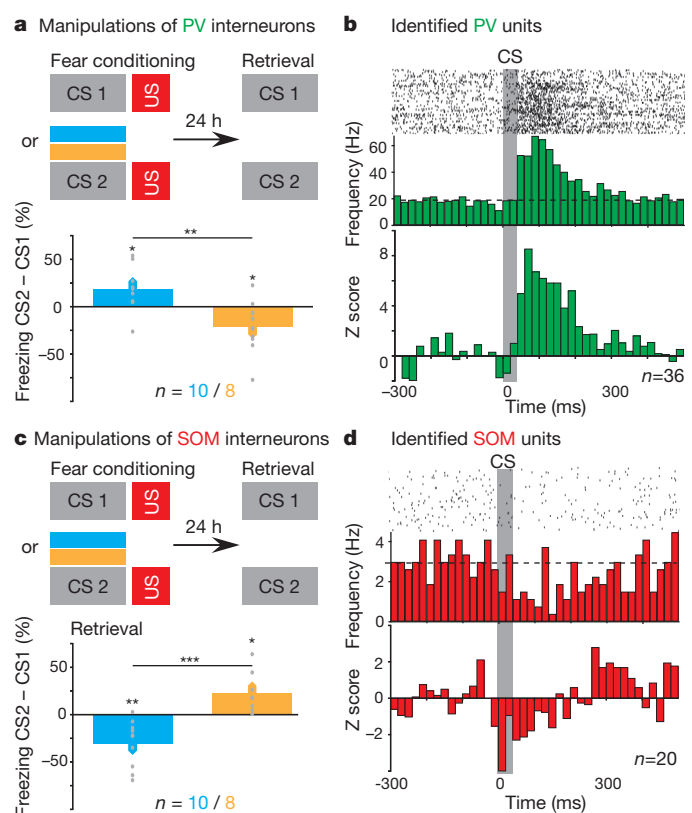


Figure 3 | Differential roles of PV⁺ and SOM⁺ interneurons during the CS. **a**, Top, conditioning as in Fig. 2a, but with optogenetic manipulation restricted to the CS. Bottom, PV⁺ activity manipulations during the CS bidirectionally affect learning, with an opposite sign compared to manipulations during the US (Fig. 2a). **b**, Optogenetically identified PV⁺ cells are excited by the CS during conditioning (top, example trace, dashed line: average baseline firing; bottom, population response). **c**, Top, animals expressing opsins in BLA SOM⁺ interneurons were conditioned as in **a**. Bottom, SOM⁺ activity manipulations affected learning bidirectionally, in the opposite way compared to PV⁺ manipulations. **d**, SOM⁺ cells are inhibited by the CS during conditioning. Values are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Statistical analysis in Methods.

(Fig. 4a, top left), we combined whole-cell recordings from interneurons ($n = 101$) in acute brain slices with optogenetic activation of PV⁺ interneurons (Fig. 4a and Extended Data Fig. 7). Post-hoc immunocytochemical identification revealed that 4 out of 11 identified SOM⁺ interneurons received reliable inhibitory input from PV⁺ cells (Fig. 4a, top right and bottom; Extended Data Fig. 7), which can be sufficient to suppress spiking in SOM⁺ interneurons (Extended Data Fig. 7). Owing to potential wash-out of the antigen SOM in whole-cell recordings, these data represent a solid lower bound rather than the actual connectivity rate (see Methods). Notwithstanding, these results clearly demonstrate the existence of a functional connection between PV⁺ and SOM⁺ interneurons in the BLA.

A prerequisite for the ability of SOM⁺ interneurons to influence directly the integration of auditory inputs in principal neurons is that they preferentially target principal neuron dendrites²⁷. To address this, we conditionally expressed green fluorescent protein (GFP) fused to the presynaptic marker synaptophysin in BLA SOM⁺ interneurons. We observed prominent synaptic contacts on the dendrites of principal neurons (Fig. 4b). In contrast, PV⁺ interneurons showed a distinct targeting pattern, preferentially contacting the perisomatic domain of principal neurons (Extended Data Fig. 8). We next asked whether SOM⁺ interneurons can control the processing of sensory information in principal neurons. Electrical stimulation of thalamic sensory afferents in acute brain slices evoked reliable excitatory postsynaptic potentials

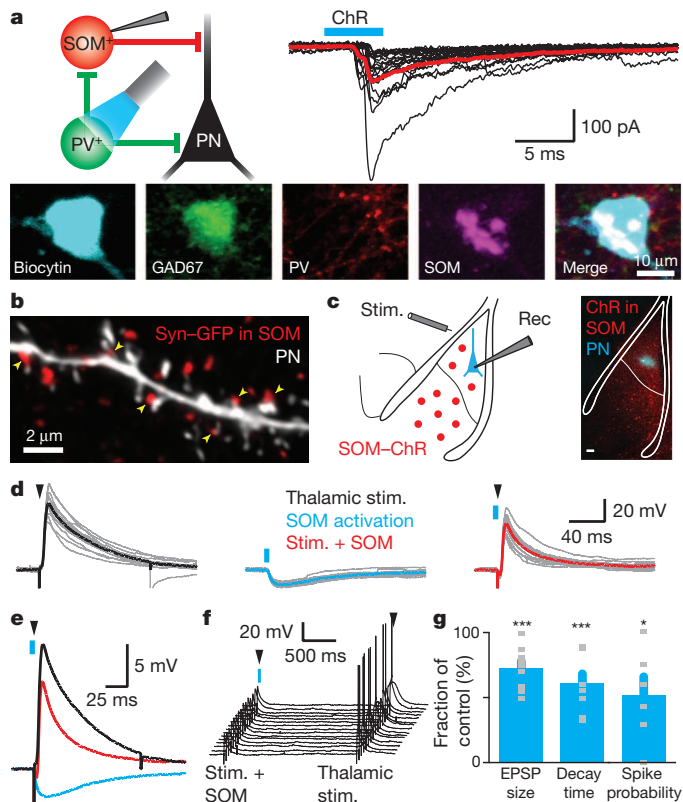


Figure 4 | PV⁺ and SOM⁺ interneurons form a disinhibitory microcircuit in the BLA. **a**, Top left, proposed PV⁺ and SOM⁺ microcircuit and experimental set-up. Top right, whole-cell recordings of EGFP⁺ cells in GAD67-EGFP::PV-Cre mice, infected with rAAV-DIO-ChR2-mCherry. Inhibitory inputs in SOM⁺ cells evoked by optogenetic activation of PV⁺ interneurons are shown. Bottom, post-hoc IHC identification of SOM⁺ cells (EGFP⁺ cells $n = 101$, of those SOM⁺ $n = 11$, SOM⁺ and input $n = 4$, see Extended Data Fig. 7c). **b**, Example of a principal neuron dendrite (white) and synaptic contacts (yellow arrowheads) from SOM⁺ interneurons (red; conditionally expressed presynaptic marker synaptophysin-GFP). **c–g**, SOM⁺ interneurons modulate excitatory inputs and neuronal output of BLA principal neurons in slices. **c**, Left, experimental design. Right, ChR2-mCherry⁺ SOM⁺ interneurons (red) and recorded principal neuron (blue) in the BLA. **d**, Example traces of excitatory thalamic (left, electrical stimulation) and inhibitory SOM⁺-cell-evoked (middle, 5 ms blue light) postsynaptic potentials as well as co-activation (right, electrical stimulation 5 ms after light onset). **e**, Averages of recordings in **d**. **f**, Spike output of principal neurons in response to thalamic afferent stimulation with and without SOM⁺ cell activation. **g**, Normalized amplitude and decay time of thalamic EPSPs and spike probability in response to thalamic EPSPs with simultaneous activation of SOM⁺ interneurons. Values are mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$. Statistical analysis in Methods.

(EPSPs)³⁸ in principal neurons (Fig. 4c–e), whereas optical stimulation of SOM⁺ interneurons evoked inhibitory postsynaptic potentials (IPSPs) (Fig. 4d, e). Co-activation caused a significant reduction of both the amplitude and the decay time of the evoked EPSPs (Fig. 4d, e, g), showing that SOM⁺ interneurons can control the integration of thalamic inputs in principal neurons (see Methods). In addition, spike output induced by supra-threshold thalamic stimulation was significantly reduced by simultaneous activation of SOM⁺ interneurons (Fig. 4f, g). These results demonstrate that PV⁺ interneurons are capable of inhibiting the output of SOM⁺ interneurons, which in turn can modulate integration of excitatory inputs in BLA principal neurons and thereby influence principal neuron spike output.

The emerging picture is that CS processing in the BLA is under the control of a microcircuit in which PV-to-SOM inhibition disinhibits principal neuron dendrites during tone presentation, while the principal neuron perisomatic domain receives increased inhibition directly

from PV⁺ interneurons. In contrast, during the US, both PV⁺ and SOM⁺ interneurons are inhibited (Fig. 2e and Extended Data Fig. 5a, b), leading to a qualitatively different pattern of inhibition along the somato-dendritic axis of principal neurons (see Discussion).

Impact on BLA network activity

To evaluate how inhibition supplied by PV⁺ and SOM⁺ interneurons affects CS processing in the BLA, we compared tone-evoked excitatory responses in putative principal neurons with and without optogenetic manipulation of the interneurons. Activation of SOM⁺ cells resulted in a strong reduction of tone-evoked activity in principal neurons, whereas inhibition of SOM⁺ interneurons caused a slight enhancement (Fig. 5a, b and Extended Data Fig. 9), supporting the idea that SOM⁺ cells directly inhibit principal neuron tone input processing and/or responses. Activation of PV⁺ interneurons caused a substantial enhancement of tone responses, whereas inhibition of PV⁺ interneurons had little effect on tone responses (Fig. 5a, b and Extended Data Fig. 9), indicating that direct PV–principal neuron inhibition is outweighed by the indirect PV–SOM–principal neuron disinhibitory pathway during CS processing (see below). In addition to effects on the average CS response of putative principal neurons, PV⁺ and SOM⁺ interneurons also differentially affected the number of tone-excited principal neurons. Whereas conditions

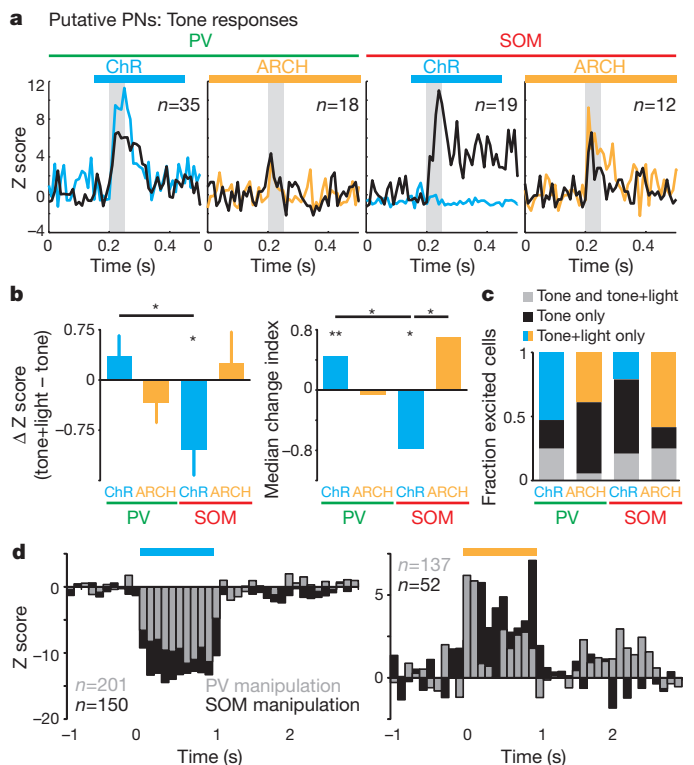


Figure 5 | Distinct effects of PV⁺ and SOM⁺ interneurons on BLA network activity. **a**, Tone-evoked population responses in tone-excited putative principal neurons with (blue or yellow) and without (black) optogenetic stimulation of PV⁺ or SOM⁺ interneurons (for selection criteria see Methods). Grey bars denote tone duration; blue and yellow bars indicate optogenetic stimulations. **b**, Left, Z-score differences in tone-excited putative principal neurons during the tone induced by optogenetic manipulations of PV⁺ or SOM⁺ interneurons. Right, median change index for the tone responses of putative principal neurons upon PV⁺ or SOM⁺ cell manipulations (see Methods). **c**, Fractions of the excited putative principal neurons, which are excited either exclusively during the tone (black), exclusively during the tone + light (blue or yellow) or in both conditions (grey). **d**, Bidirectional changes in spontaneous activity of putative principal neurons on optogenetic manipulations of PV⁺ and SOM⁺ interneurons. Activation of both interneuron types reduces (left) and inhibition enhances (right) BLA network activity. Values are mean \pm s.e.m. unless stated otherwise. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Statistical analysis in Methods.

impairing learning (SOM–Chr2 and PV–ARCH; Fig. 3 and Extended Data Fig. 2b) reduced the number of tone-excited cells, manipulations enhancing learning (PV–Chr2 and SOM–ARCH; Fig. 3 and Extended Data Fig. 2b) uncovered excitatory tone responses in large populations of putative principal neurons that were previously unresponsive (Fig. 5c and Extended Data Fig. 9). Together, these results demonstrate that PV⁺ and SOM⁺ interneurons have opposite effects on CS responses that are consistent with the observed impact on learning (Fig. 3 and Extended Data Fig. 2b)—manipulations that increased fear learning caused stronger tone responses in more principal neurons, whereas conditions that impaired learning led to smaller responses in fewer principal neurons. In addition, these data indicate that the described disinhibitory microcircuit is an important factor governing tone processing in the BLA during learning. Interestingly, spontaneous activity of putative principal neurons was not differentially affected by manipulations of the two interneuron populations (Fig. 5d and Extended Data Fig. 10a, c), suggesting that these effects are probably stimulus-specific and reflect the organization of sensory inputs onto principal neurons: auditory inputs arriving at principal neuron dendrites are affected more strongly by dendritic inhibition from SOM⁺ interneurons, potentially by locally preventing the initiation of dendritic spikes^{39,40}. In contrast, spontaneous firing is affected by both perisomatic inhibition from PV⁺ interneurons and dendritic inhibition from SOM⁺ cells.

Discussion

The level of activation of a principal neuron in the BLA during fear learning is a key factor determining whether inputs to this principal neuron will undergo synaptic plasticity and whether cellular plasticity is triggered that enables recruitment of the neuron into the memory trace^{8,41,42}. Consistent with a central role of synaptic inhibition in this process, we observed strong effects on fear memory retrieval when we manipulate distinct interneuron types during fear memory acquisition. In addition, our data point to a multifaceted role of inhibition in acquisition of conditioned fear: we find that the effects of optogenetic manipulation of interneuron activity strongly depend on both stimulus identity (CS versus US) and on interneuron type (PV⁺ versus SOM⁺). Moreover, our results identify interneuron–interneuron interactions as a novel and important factor governing auditory responses and learning.

During the presentation of an auditory cue, PV⁺ interneurons are, on average, excited, most likely through direct sensory input from auditory thalamus and auditory cortex^{7,43}. As in other cortical structures^{9,10}, this will lead to somatic feedforward inhibition in postsynaptic principal neurons, a process that is expected to restrict their activation and, as a consequence, the induction of synaptic plasticity during learning. Unexpectedly, our results indicate that increasing PV⁺ cell activity during the CS correlates with enhanced learning, and with stronger auditory responses in principal neurons. These observations can be explained in light of our finding that PV⁺ interneurons not only contact principal neurons, but also supply inhibition to dendrite-targeting SOM⁺ interneurons in the BLA. In line with this, activation of SOM⁺ interneurons reduces learning and decreases auditory responses in BLA principal neurons. Together, these results strongly suggest that during auditory processing, dendritic disinhibition in principal neurons overrides the effect of increased perisomatic inhibition, potentially through increased dendritic electrogenesis and associated burst firing (Extended Data Fig. 10b, d)²⁹ as it has been demonstrated in the hippocampus²⁶. Furthermore, it is conceivable that locally restricted regulation of dendritic inhibition adds a level of control by determining which inputs contribute to action potential firing and undergo plasticity upon association with the US²⁸.

In contrast, during the aversive US, PV⁺ cells were inhibited and optogenetic augmentation of US-induced inhibition enhanced fear learning. This indicates that inhibition of PV⁺ interneurons represents an important, permissive mechanism for fear learning, most likely by gating associative plasticity induction during the US, as has recently been shown in auditory cortex⁴⁴. Our data now extend and complement this view by demonstrating that SOM⁺ cells are also inhibited during

the US, indicating disinhibition of the entire somatodendritic axis of principal neurons, thereby strongly increasing their activity which is required for fear learning (Extended Data Fig. 6e). Moreover, because auditory inputs synapse onto dendrites^{45,46}, dendritic disinhibition represents an important, novel mechanism in auditory fear learning, favouring the induction of synaptic plasticity during the aversive US. A likely source of US-induced inhibition is other interneuron subtypes that can contact both PV⁺ and SOM⁺ interneurons^{12,44,47}, and/or fast neuromodulatory processes. An important open question pertains to the mechanisms bridging the temporal gap between CS-related firing and activity caused by the US. Potential candidates include dendritic plateau potentials and/or second messengers such as calcium^{39,40,48}, and future experiments will be required to test these possibilities.

Although we observed that most PV⁺ and SOM⁺ interneurons exhibited CS and US responses consistent with the effects of optogenetic manipulations on learning, both interneuron types also displayed response heterogeneity (Extended Data Fig. 5). This may at least partially reflect anatomically defined subtypes of interneurons, for example, PV⁺ basket or PV⁺ chandelier cells (see ref. 14). Alternatively, the functional heterogeneity might also be due to differential input or output connections of interneuron subgroups. Future advances in genetic targeting of specific neuronal subpopulations may allow a more detailed dissection of the role of these functional subgroups of interneurons.

Different types of interneurons exhibit distinct temporal dynamics during different network states that correlate with specific aspects of behaviour^{49,50}. It has been postulated that direct interactions between soma-targeting and dendrite-targeting interneurons are an important mechanism by which spatio-temporal patterns of inhibition might be regulated²⁶. Here, we provide evidence that similar mechanisms can govern associative learning and the strength of the memory trace. In the identified microcircuit, differential modulation of PV⁺ and SOM⁺ interneurons may permit flexible regulation of learning according to the behavioural context and the animal's internal state.

METHODS SUMMARY

PV–Cre³¹ or SOM–Cre³⁷ mice (2–4 months) were used. Full details of animals, materials and methods for virus injections, optogenetic stimulations, behaviour, single-unit recordings and slice electrophysiology are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 July 2013; accepted 17 March 2014.

Published online 11 May 2014.

1. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
2. Davis, M. in *The Amygdala* (ed. Aggleton, J. P.) 213–288 (Oxford Univ. Press, 2000).
3. Fanselow, M. S. & Poulos, A. M. The neuroscience of mammalian associative learning. *Annu. Rev. Psychol.* **56**, 207–234 (2005).
4. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
5. Pape, H. C. & Paré, D. Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol. Rev.* **90**, 419–463 (2010).
6. Wacław, R. R., Ehrman, L. A., Pierani, A. & Campbell, K. Developmental origin of the neuronal subtypes that comprise the amygdalar fear circuit in the mouse. *J. Neurosci.* **30**, 6944–6953 (2010).
7. Sah, P., Faber, E. S., Lopez De Armentia, M. & Power, J. The amygdaloid complex: anatomy and physiology. *Physiol. Rev.* **83**, 803–834 (2003).
8. Ehrlich, I. et al. Amygdala inhibitory circuits and the control of fear memory. *Neuron* **62**, 757–771 (2009).
9. Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5**, 793–807 (2004).
10. Somogyi, P. & Klausberger, T. Defined types of cortical interneurone structure space and spike timing in the hippocampus. *J. Physiol. (Lond.)* **562**, 9–26 (2005).
11. Freund, T. F. & Buzsáki, G. Interneurons of the hippocampus. *Hippocampus* **6**, 347–470 (1996).
12. Fishell, G. & Rudy, B. Mechanisms of inhibition within the telencephalon: “where the wild things are”. *Annu. Rev. Neurosci.* **34**, 535–567 (2011).
13. Spannato, J., Polepalli, J. & Sah, P. Interneurons in the basolateral amygdala. *Neuropharmacology* **60**, 765–773 (2011).

14. Bienvenu, T. C., Busti, D., Magill, P. J., Ferraguti, F. & Capogna, M. Cell-type-specific recruitment of amygdala interneurons to hippocampal theta rhythm and noxious stimuli *in vivo*. *Neuron* **74**, 1059–1074 (2012).
15. Heldt, S. A. & Ressler, K. J. Training-induced changes in the expression of GABA_A-associated genes in the amygdala after the acquisition and extinction of Pavlovian fear. *Eur. J. Neurosci.* **26**, 3631–3644 (2007).
16. Chhatwal, J. P., Myers, K. M., Ressler, K. J. & Davis, M. Regulation of gephyrin and GABA_A receptor binding within the amygdala after fear acquisition and extinction. *J. Neurosci.* **25**, 502–506 (2005).
17. Harris, J. A. & Westbrook, R. F. Evidence that GABA transmission mediates context-specific extinction of learned fear. *Psychopharmacology (Berl.)* **140**, 105–115 (1998).
18. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
19. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
20. Kawaguchi, Y. & Kubota, Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb. Cortex* **7**, 476–486 (1997).
21. Woodruff, A. R. & Sah, P. Networks of parvalbumin-positive interneurons in the basolateral amygdala. *J. Neurosci.* **27**, 553–563 (2007).
22. Rainnie, D. G., Mania, I., Mascagni, F. & McDonald, A. J. Physiological and morphological characterization of parvalbumin-containing interneurons of the rat basolateral amygdala. *J. Comp. Neurol.* **498**, 142–161 (2006).
23. Muller, J. F., Mascagni, F. & McDonald, A. J. Pyramidal cells of the rat basolateral amygdala: synaptology and innervation by parvalbumin-immunoreactive interneurons. *J. Comp. Neurol.* **494**, 635–650 (2006).
24. McDonald, A. J. & Betette, R. L. Parvalbumin-containing neurons in the rat basolateral amygdala: morphology and co-localization of Calbindin-D_{28k}. *Neuroscience* **102**, 413–425 (2001).
25. Muller, J. F., Mascagni, F. & McDonald, A. J. Coupled networks of parvalbumin-immunoreactive interneurons in the rat basolateral amygdala. *J. Neurosci.* **25**, 7366–7376 (2005).
26. Lovett-Barron, M. *et al.* Regulation of neuronal input transformations by tunable dendritic inhibition. *Nature Neurosci.* **15**, 423–430 (2012).
27. Muller, J. F., Mascagni, F. & McDonald, A. J. Postsynaptic targets of somatostatin-containing interneurons in the rat basolateral amygdala. *J. Comp. Neurol.* **500**, 513–529 (2007).
28. Chiu, C. Q. *et al.* Compartmentalization of GABAergic inhibition by dendritic spines. *Science* **340**, 759–762 (2013).
29. Gentet, L. J. *et al.* Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nature Neurosci.* **15**, 607–612 (2012).
30. Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. & Deisseroth, K. Circuit-breakers: optical technologies for probing neural signals and systems. *Nature Rev. Neurosci.* **8**, 577–581 (2007).
31. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
32. Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
33. Lima, S. Q., Hromadka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS ONE* **4**, e6099 (2009).
34. Woodruff, A. R. & Sah, P. Inhibition and synchronization of basal amygdala principal neuron spiking by parvalbumin-positive interneurons. *J. Neurophysiol.* **98**, 2956–2961 (2007).
35. McDonald, A. J. Neuronal organization of the lateral and basolateral amygdaloid nuclei in the rat. *J. Comp. Neurol.* **222**, 589–606 (1984).
36. Likhtik, E., Pelletier, J. G., Popescu, A. T. & Paré, D. Identification of basolateral amygdala projection cells and interneurons using extracellular recordings. *J. Neurophysiol.* **96**, 3257–3265 (2006).
37. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
38. Bissière, S., Humeau, Y. & Lüthi, A. Dopamine gates LTP induction in lateral amygdala by suppressing feedforward inhibition. *Nature Neurosci.* **6**, 587–592 (2003).
39. Humeau, Y. & Lüthi, A. Dendritic calcium spikes induce bi-directional synaptic plasticity in the lateral amygdala. *Neuropharmacology* **52**, 234–243 (2007).
40. Murayama, M. *et al.* Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* **457**, 1137–1141 (2009).
41. Han, J. H. *et al.* Neuronal competition and selection during memory formation. *Science* **316**, 457–460 (2007).
42. Zhou, Y. *et al.* CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nature Neurosci.* **12**, 1438–1443 (2009).
43. Woodson, W., Farb, C. R. & Ledoux, J. E. Afferents from the auditory thalamus synapse on inhibitory interneurons in the lateral nucleus of the amygdala. *Synapse* **38**, 124–137 (2000).
44. Letzkus, J. J. *et al.* A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480**, 331–335 (2011).
45. Farb, C. R. & LeDoux, J. E. NMDA and AMPA receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse* **27**, 106–121 (1997).
46. McDonald, A. J. Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.* **55**, 257–332 (1998).
47. Jiang, X., Wang, G., Lee, A. J., Stornetta, R. L. & Zhu, J. J. The organization of two new cortical interneuronal circuits. *Nature Neurosci.* **16**, 210–218 (2013).
48. Lisman, J., Yasuda, R. & Raghavachari, S. Mechanisms of CaMKII action in long-term potentiation. *Nature Rev. Neurosci.* **13**, 169–182 (2012).
49. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321**, 53–57 (2008).
50. Kvitsiani, D. *et al.* Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).

Acknowledgements We thank all members of the Lüthi laboratory and E. Schuman for discussions and critical comments on the manuscript. We would like to thank J. Lüdke, P. Argast and P. Buchmann for technical assistance. We further thank the Facility for Imaging and Microscopy at the FMI, in particular S. Bourke and M. Kirschmann, for their assistance in image acquisition and analysis. We also thank K. Deisseroth, E. Boyden, J. Huang, R. Sprengel and S. Arber for sharing materials and mouse lines. This work was supported by the Novartis Research Foundation, by the National Center of Competences in Research: ‘SYNAPSY — The Synaptic Bases of Mental Diseases’ (financed by the Swiss National Science Foundation) as well as by a SNSF core grant to A.L. S.B.E.W. is supported by a Schering Foundation Fellowship. J.J.L. is supported by a Swiss National Science Foundation Ambizione Fellowship. J.G. and G.A.J. are supported by EMBO Long-Term Fellowships and Marie Curie Action Fellowships.

Author Contributions S.B.E.W. initiated the project, performed most experiments and data analysis, and wrote the manuscript. J.G. performed and analysed *in vitro* experiments and immunohistochemistry. P.T. performed and analysed immunohistochemistry. S.K. performed *in vitro* experiments and analysed subcellular targeting experiments. G.A.J. and R.W.F. performed and helped with data analysis. C.M. provided technical assistance. C.H. and I.E. helped to establish optogenetic manipulations and single-unit recordings. S.B.E.W., J.J.L. and A.L. conceived the project and wrote the manuscript. All authors contributed to the experimental design and interpretation, and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. (andreas.luthi@fmi.ch) or J.J.L. (johannes.letzkus@brain.mpg.de).

METHODS

Animals. Male PV-Cre (ref. 31) or SOM-Cre (ref. 37) mice (2–4 months) were individually housed in a 12 h light/dark cycle while the experiments were running. Before behavioural experiments, animals were habituated to the experimenter by handling for ≥ 3 times. Food and water were available ad libitum. Behavioural experiments were performed during the light cycle. All animal procedures were performed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Virus injection and optogenetics. Mice were anaesthetized with isoflurane (Attane, Provect; induction 5%, maintenance 1.5%) in oxygen-enriched air (Oxymat 3, Weimann) and fixed in a stereotactic frame (Kopf Instruments). Local injections of ropivacain (Naropin, AstraZeneca) under the scalp and systemic injections of meloxicam (60 μ l of 0.5 mg ml⁻¹, intraperitoneal, Metacam, Boehringer Ingelheim) were provided for analgesia. A feedback-controlled heating pad (FHC) assured maintenance of the core body temperature at 36.5 °C.

For selective expression of opsins, conditional Cre-dependent recombinant adeno-associated viruses (rAAV) were injected into the basolateral amygdala (BLA) of PV-Cre or SOM-Cre mice, respectively. For expression of the excitatory channelrhodopsin-2 (ref. 30) we used a custom rAAV 2/7 EF1a::DIO-ChR2(H134R)-2A-NpHR-2A-Venus^{44,51} or rAAV 2/5 EF1a::DIO-ChR2(H134R)-mCherry, and for expression of the inhibitory proton pump ARCH³², we used an rAAV 2/5 CBA::flex-ARCH-GFP. For specific expression of ARCH in principal neurons, we used an rAAV 2/5 CaMKII::ARCH-GFP (all rAAVs Vector Core, University of Pennsylvania). rAAVs (~0.5 μ l per hemisphere) were delivered into the BLA using glass pipettes (tip diameter 10–20 μ m) connected to a picospritzer (Parker Hannifin Corporation) at the following coordinates: 1.7 mm posterior to bregma, ± 3.3 mm lateral to the midline and 4.1–4.2 mm below the cortical surface. For optogenetic manipulations during behaviour without electrophysiological recordings, animals were additionally implanted with bilateral custom-built optic fibre connectors (fibre: 0.48 numerical aperture, 200 μ m diameter, Thorlabs). Fibre tips were lowered to 300 μ m above the BLA at the injection site (Extended Data Fig. 1a). Implants were fixed to the skull with skull screws, cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus). After surgery, animals were allowed to recover for at least 4 weeks before behavioural training to ensure sufficient expression of the opsins. For optogenetic manipulations during behaviour, the implanted fibres were connected to a custom-built laser bench using an AOTF (AA Opto-Electronic) to control laser intensity (lasers: MBL473, 473 nm wavelength and MGL593.5, 593.5 nm wavelength, CNi Lasers). To ensure that animals could move freely, the connecting fibres were suspended over the behavioural context.

Optrode recordings and optogenetic identification of interneurons. In most animals one optical connector was replaced by a custom-built optrode to allow simultaneous optogenetic stimulations and single-unit recordings during behaviour. Optrodes consisted of an optic fibre connector with an electrode attached directly to the fibre, with the tip protruding approximately 300–500 μ m beyond the fibre. Electrodes were made of 16 individually insulated, gold-plated nichrome wires (13 μ m inner diameter, impedance 30–100 k Ω , Sandvik), attached to a connector (18 pin, Omnetics)⁴⁴. The whole implant was fixed to the skull as described above. Single-unit recordings were performed as described below.

For optogenetic identification of PV⁺ or SOM⁺ interneurons, we tested all recorded units in all animals for light responsiveness³³. We presented pulses of either blue or yellow light to activate ChR2 or ARCH, respectively (300 ms, 60 pulses, 2-s inter pulse interval, 15–20 mW at fibre tip). Units were considered as light responsive if they showed significant, time-locked (<7 ms) changes in neuronal activity on illumination (Fig. 2c and Extended Data Fig. 3i). The onset of optogenetic excitation was determined as the average latency of the first light-evoked spike in each trial. To determine the onset of inhibition reliably in neurons with low baseline firing rates, we applied a change-point analysis⁵² (Change Point Analyzer 2.0, Taylor Enterprises Inc.). Change point analysis identifies the time point exhibiting a significant change in neuronal activity relative to the preceding time points. Change points are graphically represented by a change in the slope of a plot showing the cumulative sums of the firing frequency averaged over the 60 light trials (Extended Data Fig. 3e).

To ensure that light-evoked and spontaneous spikes of identified interneurons were indeed originating from the same neuron, we calculated linear correlation (r) values for spontaneous and light-evoked spikes to quantitatively determine the similarity of their waveform shapes. Only groups of spikes with $r > 0.95$ were considered as originating from the same cell (Extended Data Fig. 3d, j).

Behaviour and optogenetic manipulation of behaviour. Before behavioural training, light-responsive interneurons were identified in opsin-expressing PV-Cre or SOM-Cre animals, as described above (Fig. 2b–d and Extended Data Fig. 3). Depending on the presence of identified interneurons, mice were submitted to different auditory fear-conditioning paradigms. To determine the physiological activity of identified interneurons during fear learning, we used a discriminative fear-conditioning paradigm without optogenetic manipulations⁵³. On training day 1

(habituation), animals were presented with 4 repetitions of a CS⁺ and a CS[−] (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 7.5 kHz or white noise (WN), counterbalanced for randomization, 75 dB sound pressure level). On day 2 (fear conditioning) animals were conditioned by pairing the CS⁺ with the US (1-s footshock, 0.75 mA DC, 5 CS/US pairings; inter-trial interval: 20–180 s; US applied at the time of next expected pip occurrence, ~1.1 s after last pip). The CS[−] was intermingled with the CS⁺, but never reinforced by a US. During fear retrieval on the next day, 4 repetitions of first the CS[−] and then of the CS⁺ were presented to test for fear learning. Single-unit activity was recorded during all behavioural sessions. Light responsiveness of the identified units was tested before habituation and after the last behavioural session.

Animals without identified interneurons were subjected to a fear conditioning paradigm combined with optogenetic manipulations. In animals with implanted optrodes, single-unit activity of non-light-responsive units was recorded during all behavioural sessions. On day 1 (habituation), two different CSs (CS1 and CS2) were presented 4 times to the animal (total CS duration of 10 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 3 kHz or 12 kHz, counterbalanced for randomization, 75 dB sound pressure level). In addition, animals were presented with 4 repetitions of the light stimuli, which were used in the subsequent conditioning, without a CS. These stimulations were also used to determine baseline effects on freezing and locomotion. On day 2 (fear conditioning), animals were fear conditioned by pairing both CS1 and CS2 with a US (1-s foot shock, 0.75 mA DC, 1 CS–US pairing each, interval between CSs: 180 s; US applied at the time of next expected pip occurrence, ~1.1 s after last pip). While one of the two CSs served as an internal control, the other CS was paired with light stimulation. Light was applied either only during the CS (blue light: 300-ms pulses starting 50 ms before each pip; yellow light: continuous illumination for 10 s starting with the onset of first CS pip), only during the US (blue or yellow light: 1-s continuous illumination, coinciding with the US), or during CS and US (combination of the described stimulation patterns). The order of CS1 and CS2 and the pairing with light stimulation were counterbalanced within behavioural groups. On day 3 (retrieval), animals were tested for fear memory retrieval. Both CS1 and CS2 were presented 4 times each without light stimulation or reinforcement. Freezing induced by the control CS and the light-paired CS was compared to determine the effect of the optogenetic manipulation in each animal (Extended Data Fig. 2). The order of CS1 and CS2 during the test was counterbalanced within behavioural groups. Because the behavioural effects of optogenetic manipulations could not be pre-specified before the experiments, we chose sample sizes with enough repetitions to allow for identification of outliers and for validation of experimental reproducibility ($n = 7–11$).

To test for the effects of optogenetic manipulations of interneuron activity on footshock- or tone-induced activity in the BLA network, we performed additional behavioural sessions in animals with implanted optrodes. We analysed footshock and tone responses in unidentified cells. This population consists mostly of principal cells with a minority of interneurons^{5,7,12,35,36}. To exclude fast-spiking interneurons, we restricted the analysis to cells with a baseline firing rate <3 Hz. To determine effects on footshock-induced activity, we delivered 20 footshocks (1 s, 0.75 mA DC) to the animals, alternately with and without coincident stimulation with blue or yellow light. To study the effects of interneuron activity manipulations on tone processing, we used 8 presentations of tones (white noise, 30 s total duration, consisting of 50-ms pips repeated at 0.9 Hz, 75 dB sound pressure level) alternately with and without coincident light stimulation. For activation of ChR2, we used a train of blue light stimuli (300-ms pulses starting 50 ms before each pip) and for ARCH we used continuous illumination with yellow light (30 s), as for the manipulations during fear learning (see above). For the analysis of the light-induced changes in shock or tone responses see the corresponding sections below.

Fear conditioning was performed in a context (context A) different from that used for all other behavioural sessions (context B). Context A was cleaned with 70% ethanol and context B with 1% acetic acid. Freezing behaviour was quantified using an automatic infrared beam detection system (Coulbourn Instruments) as described previously⁵³. The animals were considered freezing if no movement was detected for 2 s.

In vivo electrophysiology and analysis. Single-unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) as previously described⁵³. In summary, principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces. Clusters containing similar waveforms were manually defined (Extended Data Fig. 1b–e). A group of waveforms was considered to be originating from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units (determined by MANOVA) and if it displayed a clear refractory period (>1 ms) in the auto-correlogram. In addition, we used two parameters to quantify the overall separation between identified clusters in an individual channel. The J3 statistic corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin

validity index (DB) reflects the ratio of the sum of within-cluster scatter to between-cluster separation. To obtain control values for these statistics, two clusters from the central cloud of points in the principal component space from channels without detectable units were artificially defined. High values for the J3 and low values for the DB compared to the control values indicate good single-unit isolation (Extended Data Fig. 1c). Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analysed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioural training were considered. Long-term single-unit stability was evaluated using Wavetracker software (Plexon) to calculate principal component space-cylinders from a 2-min recording of spontaneous activity before each training session. Straight cylinders suggest stability of the unit isolation during all behavioural sessions (Extended Data Fig. 1d). In addition, we determined single-unit stability by quantitative evaluation of the similarity of waveform shapes by calculating linear correlation (r) values between average waveforms obtained over training days⁵³. As a control, we computed the r values from average waveforms of different neurons (Extended Data Fig. 1e).

To exclude multiple recordings of the same neuron on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at the time when the reference neuron fired, only one of the two neurons was considered for further analysis.

Neuronal activity is reported either as frequency or as Z-score value. Z-score values were calculated by generating a peri-stimulus time histogram (PSTH) of the binned neuronal data and summation of all spikes of all cells in each individual bin. The average baseline count preceding stimulus onset was subtracted from this population spike count. This difference was divided by the baseline standard deviation to obtain the Z score.

To analyse bursting activity we used the Poisson surprise method⁵⁴. Bursts were compared between baseline activity and illumination periods (blue or yellow light; 1-s illumination and 1-s baseline before light onset). Briefly, this method detects bursts by finding at least 3 spikes with consecutive interspike intervals (ISI) shorter than half the mean ISI, and calculates the 'unlikelyness' or 'surprise' that a given ISI would be expected if the spike train was generated by a Poisson process. The Poisson surprise method is a rigorous detector of bursts, because it is insensitive to fluctuations in average firing rate. We chose a surprise value of 3, corresponding to a probability of 0.05 of finding a burst of spikes within a random sequence.

Analysis of optogenetic modulation of shock responses in putative principal neurons. Shock responses and their modulation by optogenetic activation or silencing of PV⁺ cells were studied in putative principal neurons as described above. Analysis was restricted to putative principal neurons, defined as units whose mean firing rate during a 100-s baseline period was <3 Hz (ref. 36). A peri-stimulus time histogram for the shock stimulus, $PSTH^S(t)$, was calculated with 50-ms resolution between $t = -10$ s to $t = 2$ s relative to shock onset for each cell. From this the mean (μ_B^S) and standard deviation (σ_B^S) of the firing rate preceding each shock ($t = -10$ s to $t = 0$ s) were calculated and used to generate the Z-score normalized shock-triggered PSTH:

$$Z^S(t) = \frac{PSTH^S(t) - \mu_B^S}{\sigma_B^S}$$

Units were defined as shock responsive (S) if the response in one of the 20 time bins between $t = 0$ s and $t = 1$ s exceeded a Z value of 3.72 (corresponding to $P < 0.001$, corrected for multiple comparisons). Analysis was performed for the shock responses in the presence of light (SL) in the same way and all units showing excitatory shock responses in at least one of the two conditions were pooled. Response magnitudes for the S and SL conditions were defined as:

$$R^S = Z^S(t)_{t=0-1\text{ s}}$$

$$R^{SL} = Z^{SL}(t)_{t=0-1\text{ s}}$$

correspondingly. To quantify the effect of light on the shock responses, a change index (CI) was calculated for each unit:

$$CI = \frac{R^{SL} - R^S}{R^{SL} + R^S}$$

For purely excitatory responses, CI values can range between +1 (units that respond to shock only in the presence of light) and -1 (units that respond to shock only in the absence of light). CI values outside the range (-1,1) indicate inhibitory responses to one of the conditions. A value of 0 indicates no effect of light. For both animal groups (PV-ChR2, PV-ARCH), several measures were used to quantify the effect of light on shock responses at the population level. First, the average differences of the responses ($R^{SL} - R^S$) were calculated (Fig. 2g, left). Next, the median CI for the shock responsive units in each group was calculated (Fig. 2g, right). Finally,

the fraction of S only, SL only, and S and SL units from all responsive units was also calculated for both animal groups (Fig. 2h). To obtain the population activity of all shock-excited units, all spikes of all excited units were summed in each individual bin of the PSTH. From this the mean and standard deviation of the population baseline ($t = -10$ s to $t = 0$ s) were determined and used to generate the Z-score normalized shock-triggered PSTH for the population of shock-excited units (Fig. 2f).

Analysis of optogenetic modulation of tone responses in putative principal neurons. Tone responses and their modulation by optogenetic activation or silencing of PV⁺ or SOM⁺ cells were studied in putative principal neurons as described above. Tones were 50-ms white noise bursts presented at a rate of 0.9 Hz. Each block contained 27 tone pips, but only the first 9 were analysed, to match the behavioural paradigm.

Analysis was restricted to putative principal neurons, defined as units whose mean firing rate during a 100-s baseline period was <3 Hz (ref. 36). A peri-stimulus time histogram for the tone stimulus, $PSTH^T(t)$, was calculated with 10-ms resolution between $t = -200$ ms to $t = 300$ ms relative to tone onset. The mean (μ_B^T) and standard deviation (σ_B^T) of the firing rate preceding each tone pip ($t = -200$ ms to $t = 0$ ms) were calculated and used to generate Z-score normalized tone-triggered PSTH:

$$Z^T(t) = \frac{PSTH^T(t) - \mu_B^T}{\sigma_B^T}$$

Units were defined as tone responsive (T) if the response in one of the 6 time bins between $t = 0$ ms and $t = 60$ ms exceeded a Z value of 3.59 (corresponding to $P < 0.001$, corrected for multiple comparisons). Analysis was similarly performed for the tone responses in the presence of light. Cells were defined as responsive to tone in the presence of light (TL) in a similar fashion, the only modification being that for ChR2 stimulation, which was non-continuous and consisted of 300-ms light pulses preceding the tone by 50 ms, only the time period between $t = -50$ ms and $t = 0$ ms was used for calculating μ_B^{TL} and σ_B^{TL} . This was done to avoid confounding tone responses with firing rate changes induced indirectly by the light, which could change the baseline and as a result increase the standard deviation estimate. All units showing excitatory tone responses in at least one of the two conditions were pooled. Response magnitudes for the T and TL conditions were defined as:

$$R^T = Z^T(t)_{t=0-60\text{ ms}}$$

$$R^{TL} = Z^{TL}(t)_{t=0-60\text{ ms}}$$

correspondingly. To quantify the effect of light on the tone responses, a change index (CI) was calculated for each unit:

$$CI = \frac{R^{TL} - R^T}{R^{TL} + R^T}$$

For purely excitatory responses, CI values can range between +1 (units that respond to tone only in the presence of light) and -1 (units that respond to tone only in the absence of light). CI values outside the range (-1,1) indicate inhibitory responses to one of the conditions. A value of 0 indicates no effect of light. For each animal group (PV-ChR2, PV-ARCH, SOM-ChR2, SOM-ARCH), several measures were used to quantify the effect of light on tone responses at the population level. First, the average differences of the responses ($R^{TL} - R^T$) were calculated (Fig. 5b, left). Next, the median CI for the tone-responsive units in each group was calculated (Fig. 5b, right). Finally, the fraction of T only, TL only, and T and TL units from all responsive units was also calculated for each animal group (Fig. 5c). To obtain the population activity of all tone-excited units, all spikes of all excited units were summed in each individual bin of the PSTH. From this the mean and standard deviation of the population baseline ($t = -200$ ms to $t = 0$ ms) were determined and used to generate the Z-score normalized tone-triggered PSTH for the population of tone-excited units for each animal group (Fig. 5a).

Slice electrophysiology. Coronal brain slices were prepared from 3–6-month-old PV-Cre::GAD-EGFP or SOM-Cre::GAD-EGFP mice at least 4 weeks after virus injection. Animals were anaesthetized with isoflurane and decapitated in ice-cold artificial cerebrospinal fluid (ACSF in mM: NaCl 124, NaH₂PO₄ 1.25, KCl 2.7, NaHCO₃ 26, ascorbic acid 2.25, CaCl₂ 2, MgCl₂ 8.7, glucose 18). 300- μ m-thick amygdala slices were cut with a sapphire blade (vibrating microtome: Microm HM 650 V), transferred to an interface chamber and incubated for 45 min at 37 °C in ACSF (1.3 mM MgCl₂). Slices were next transferred to a recording chamber on an upright microscope (Olympus, BX61W1) and superfused with ACSF (as above, except for 1.3 mM MgCl₂, 2.5 mM CaCl₂, 10 μ M CNQX, 10 μ M CPP, 35 ± 1 °C). Interneurons were identified as EGFP-positive, mCherry-negative cells using epifluorescence (Lumen Dynamics, X-Cite 120 Hg-Lamp, Q-Imaging, RETIGA EXi camera, Extended Data Fig. 7a). Electrophysiological recordings from basolateral amygdala interneurons and pyramidal cells were acquired in whole-cell mode using 3–5 M Ω glass pipettes. Internal solutions were as follows: voltage clamp (in mM): CsCl 110,

K-gluconate 30, EGTA 1.1, HEPES 10, CaCl₂ 0.1, Mg-ATP 4, Na-GTP 0.3, 0.4% biocytin, QX 314 chloride 4; current clamp (in mM): K-methylsulphate 130, KCl 5, HEPES 10, EGTA 0.6, Na-phosphocreatine 10, Mg-ATP 4, Na-GTP 0.3, 0.4% biocytin. ChR2-expressing SOM⁺ or PV⁺ interneurons were optogenetically stimulated with a blue light LED (465 nm, Plexon, LED-driver LD-1) or a blue light laser (473 nm, see above) coupled to an optical fibre, which was positioned above the slice (stimulation intensities: 10–20 mW). Importantly, connectivity experiments (Fig. 4a and Extended Data Fig. 7) were designed to provide positive evidence for a PV–SOM connection, but for technical reasons did not allow to determine precise connectivity rates. The experimental approach we have chosen essentially excludes the detection of false-positive cells, and therefore provides strong evidence for the existence of this connection. In contrast, it is likely that false-negative results have lowered the actual number of SOM⁺ cells receiving PV⁺ cell input in our experiments. Because IPSCs were measured in the whole-cell configuration, wash-out of the antigen SOM during the recording is likely to prevent correct post-hoc identification of SOM⁺ cells in some cases. Consistent with this, we see an under-representation of SOM⁺ interneurons in our sample, compared to previous studies⁵⁵. Therefore, based on our experiments, we cannot determine the actual connectivity rate between PV⁺ and SOM⁺ interneurons, but rather obtain a solid lower bound for this connectivity.

Excitatory postsynaptic potentials (EPSPs) were elicited with a bipolar stimulation electrode (twisted platinum/iridium wire) placed in the slice medial to the LA to stimulate thalamic afferents. This approach allowed us to partially mimic the *in vivo* situation and to determine the impact of the population of SOM⁺ interneurons on the integration of thalamic inputs and CS processing in principal neurons. For spike probability experiments, stimulus intensity was adjusted to yield EPSPs close to spike threshold, such that control EPSP-evoked spike probability was less than 1. Current clamp recordings for synaptically stimulated action potentials were performed without excitatory neurotransmitter blockers and V_m was set at –60 mV via somatic current injection. Data were acquired with a Multiclamp 700B amplifier, Digidata 1322A A/D converter and pClamp 9 software (all Axon instruments) at 50 kHz and filtered at 4 or 10 kHz in voltage or current-clamp mode, respectively. Series resistance (Rs) was monitored continuously and experiments were typically discarded when Rs exceeded 30 M Ω . Outside-out patches were pulled at the end of the experiment and cells were fixed for 1 h in 4% paraformaldehyde before immunohistochemistry.

Determination of the subcellular distribution of synaptic contacts. To determine the distribution of synaptic contacts made by SOM⁺ or PV⁺ interneurons on different subcellular compartments of principal neurons, we injected a rAAV for the conditional expression of GFP, coupled to the presynaptic marker synaptophysin⁵⁶ into the BLA of 3-month-old SOM–Cre or PV–Cre animals, respectively. Two to four weeks after virus injection, coronal brain slices for slice electrophysiology were prepared as explained above. Using the voltage-clamp internal solution, pyramidal neurons in the basolateral amygdala (one cell per slice) were filled with biocytin for 10–15 min before pulling an outside-out patch. Brain slices were fixed for 1 h in 4% paraformaldehyde and stored in PBS. Incubation with streptavidin-conjugated Alexa 568 (1:1,000, Invitrogen) for visualization of the biocytin-filled pyramidal neuron was done overnight at 4 °C. No staining for synaptophysin-coupled GFP was performed because native fluorescence was sufficient.

Imaging was carried out using a confocal laser scanning microscope (LSM700, Zeiss) equipped with 488 nm and 555 nm laser diodes. Overview images of biocytin-filled neurons were acquired with a 20 \times objective (2.5 μ m z-sections). Soma and dendritic branches were scanned with a 63 \times objective (Plan-Apochromat 63 \times /1.40 Oil DIC, Zeiss; image size 1,012 \times 1,024 pixels, pinhole 1 Airy unit, 2.3-fold digital zoom, 130 nm z-interval). Sections from both apical and basal dendrites were chosen at a distance of 50 μ m, 100 μ m and 200 μ m from the soma.

Deconvolution of 3D images was performed using Huygens Software (Scientific Volume Imaging). Contact points between presynaptic boutons as defined by synaptophysin–GFP expression and postsynaptic principal neuron compartments were identified in deconvolved 3D images and manually counted using Imaris Software (Bitplane AG). The length of individual dendritic sections was determined in 3D using the TREES toolbox (<http://www.treestoolbox.org>, ref. 57) in Matlab (Mathworks). Total number of dendritic contacts from SOM⁺ and PV⁺ interneurons was extrapolated assuming a total length of the dendritic tree of 2,000 μ m⁵⁸.

Histological verification of optrode placement and virus expression. After completion of experiments, mice were deeply anaesthetized with avertin (0.3 g kg^{–1}) and an electrolytic lesion was made at the electrode tip by applying 0.1 μ A for 15 s to two of the electrode wires. Mice were then transcardially perfused with phosphate-buffered saline (PBS) followed by freshly made, ice-cold 4% paraformaldehyde (PFA). Brains were extracted and kept in PFA overnight for post-fixation. Coronal, 60- μ m-thick brain slices were then cut with a vibratome (VT1000 S, Leica) and stored in PBS containing 0.05% sodium azide. To visualize virus expression, standard immunostaining procedures were performed on free-floating brain sections: overnight incubation at 4 °C to chicken anti-GFP antibody (1:1,000, catalogue no.

A10262, Invitrogen), together with mouse anti-NeuN (MAB377, Invitrogen), 2 h incubation with anti-chicken Alexa 488, and anti-mouse Alexa 594 antibodies, as well as 5 min exposure to 4',6-diamidin-2-phenylindol (DAPI, 1:10,000) as a counterstain. For PV and SOM staining, mice were transcardially perfused with 4% PFA additionally containing 1.5% picric acid and 0.05% glutaraldehyde, and post-fixation was omitted. Before application of the primary antibody, slices were incubated for 30 min at room temperature with iFX-enhancer (Invitrogen) to achieve a better signal-to-noise ratio. Slices were then incubated at 4 °C for 48 h with either rat anti-SOM antibody (1:500, catalogue no. MAB354, Millipore) or anti-PV antibody raised in guinea-pig (1:500, catalogue no. 195004, Synaptic Systems), followed by overnight incubation at 4 °C with secondary anti-rat or anti-guinea-pig Alexa 647 antibodies (1:1,000, both Invitrogen). Slices were mounted on gelatin-coated glass slides, coverslipped and imaged using a stereomicroscope (Leica) with a monochrome camera (Qimaging) and LED (Lumencor) light source. Using the merged epifluorescence images of virus expression and counterstain as well as the reflected light image of the same sections, we determined viral expression as well as placement of the electrode and optical fibres by assessing the electrolytic and mechanical lesion produced by the electrode tip or fibre, respectively. Mice were included into the analysis if they showed virus expression bilaterally within BLA, the electrode tip was placed inside of the BLA and fibre tip placement was not more than approximately 400 μ m away from BLA. Inclusion of animals was done blind to the behavioural group and the behavioural results.

Histological verification of specificity of viral expression. To test specificity of viral delivery of opsins, we quantified co-localization of the virally transduced opsin tag with the corresponding peptide antigen visualized by immunohistochemistry (IHC). Therefore, we stained 60- μ m-thick brain sections from perfused and picric acid/PFA fixed brains for GFP (chicken anti-GFP, A10262, Invitrogen, 1:1,000) and either PV (guinea-pig anti-PV, 195004, Synaptic Systems, 1:500) or SOM (rat anti-SOM, MAB354, Millipore, 1:500) using protocols as described above. Per mouse ($n = 2$ for each genotype, PV and SOM), three slices of the amygdala corresponding to anterograde, medial and posterior regions ($n = 6$ per genotype), respectively, of the BLA were scanned using a confocal microscope (20 \times objective, z-stacks of 1 μ m). Image stacks were 3D-cropped to 400 μ m \times 400 μ m \times stack height (21–40 μ m) for anterograde amygdala, and 500 μ m \times 500 μ m \times stack height (20–48 μ m) for medial and posterior amygdala regions of interest. Within these regions, cells positive for the viral tag and/or the IHC label were manually counted (Imaris software, Bitplane AG), cell numbers normalized, averaged, and co-localization calculated.

Histological verification of cell identity in patched slices. After completion of *in vitro* electrophysiological experiments, 300- μ m-thick slices were post-fixed for 1 h in fresh 4% PFA, then washed and stored in PBS. Staining for GFP, PV and SOM was performed as stated above, and for visualization of the biotin-filled patched cell, streptavidin conjugated to Alexa 405 (Invitrogen) was added together with the secondary antibodies. Staining for the mCherry tag of the injected virus was omitted due to sufficient fluorescence of the native protein. Slices were imaged using a confocal laser scanning microscope (LSM700, LSM710; Zeiss) to determine quadruple co-localization of PV or SOM, respectively, with GFP, mCherry and Biotin. Notably, we cannot exclude wash-out of the stained antigens during the whole-cell patch-clamp recordings. Therefore, our approach, rather than providing a precise estimation of connectivity rates between PV⁺ and SOM⁺ interneurons, provides a solid lower bound of those connections.

Statistical analysis. All distributions passed tests for normality (Kolmogorov–Smirnov) and for equal variance (Levene Median), unless noted differently. Statistical tests were performed for specific experiments as follows.

Figure 1d: effects of optogenetic PV⁺ activation during CS and US on acquisition of fear memory. Comparison of freezing to two different CSs (internal control and with PV⁺ cell manipulation during conditioning) in experimental and control animals ($n = 11$ and $n = 10$, respectively). Two-way ANOVA identified a significant interaction between group and CS: $F_{(1,38)} = 4.824$, $P < 0.05$. Post-hoc pairwise Tukey's tests revealed significant differences between CS1 and CS2 ($P < 0.05$), between CS1 and CS2 within the experimental group ($P < 0.01$) and between the control and the experimental group within CS2 ($P < 0.05$).

Figure 2a: comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for PV⁺ cell activation ($n = 11$) or inhibition ($n = 9$) during the US. Two-tailed, unpaired Student's *t*-test revealed a significant difference ($P < 0.001$). In addition, freezing levels to the two CSs were compared separately for PV⁺ cell activation and inhibition. Two-tailed, paired Student's *t*-tests revealed significant differences ($P < 0.05$ for PV⁺ cell activation; $P < 0.001$ for PV⁺ cell inhibition).

Figure 2d: comparisons of spike half-width and spontaneous activity between optogenetically identified PV⁺ cells ($n = 36$) and putative principal neurons from the same recordings ($n = 142$). Owing to a non-normal distribution, the Mann–Whitney rank sum test was used and revealed significant differences (for both: $P < 0.001$).

Figure 2g: left: differences in average Z scores of putative principal neurons during the footshock with and without optogenetic activation ($n = 56$) or inhibition ($n = 56$) of PV⁺ interneurons. Owing to non-normal distributions, the Mann–Whitney rank sum test was used and revealed a significant difference between the effects of PV⁺ activation and inhibition ($P < 0.001$). In addition, the Wilcoxon signed rank test was used to determine statistical significance of the differences in each individual group and revealed significant differences between the footshock responses with and without light for both PV⁺ interneuron manipulations (PV⁺ interneuron activation $P < 0.01$; PV⁺ interneuron inhibition $P < 0.05$). Right: no significant differences were found for the median change index in putative principal neurons for PV⁺ interneuron activity manipulations.

Figure 3a: Comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for PV⁺ cell activation ($n = 10$) or inhibition ($n = 8$) during the CS. Two-tailed, unpaired Student's *t*-test revealed a significant difference ($P < 0.01$). In addition, freezing levels to the two CSs were compared separately for PV⁺ cell activation and inhibition. Two-tailed, paired Student's *t*-tests revealed significant differences ($P < 0.05$ for PV⁺ cell activation and inhibition).

Figure 3c: comparison of differences between CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for SOM⁺ cell activation ($n = 10$) or inhibition ($n = 8$) during the CS. Two-tailed, unpaired Student's *t*-test revealed a significant difference ($P < 0.001$). In addition, freezing levels to the two CSs were compared separately for SOM⁺ cell activation and inhibition. Two-tailed, paired Student's *t*-tests revealed significant differences ($P < 0.01$ for SOM⁺ cell activation and $P < 0.05$ for SOM⁺ cell inhibition).

Figure 4g: comparisons of EPSP size ($n = 9$ cells), EPSP decay time ($n = 9$) and of spike probability ($n = 6$) between electrical stimulation and electrical stimulation with optogenetic activation of SOM⁺ cells. Data were normalized to control conditions. Two-tailed, paired Student's *t*-tests revealed significant differences ($P < 0.05$ for spike probability; $P < 0.001$ for EPSP size and decay time).

Figure 5b: left, comparisons of the differences of average tone responses with and without optogenetic manipulations between the different animal groups (PV–ChR2, $n = 36$ cells; PV–ARCH, $n = 18$; SOM–ChR2, $n = 19$; SOM–ARCH, $n = 12$). One-way ANOVA indicated significant main effects between animal groups: $F_{(3,80)} = 3.166$, $P < 0.05$. Post-hoc pairwise Tukey's tests revealed a significant difference between PV–ChR2 and SOM–ChR2 ($P < 0.05$). In addition, two-tailed, paired Student's *t*-test was used to determine statistical significance of the differences in each individual group and revealed an effect in the SOM–ChR2 group ($P < 0.05$). Right, comparisons of the median change index for the tone responses with and without light between the different animal groups (as in the left panel). Owing to a non-normal distribution, the Kruskal–Wallis one-way ANOVA on ranks test was used and revealed significant differences between the different animal groups ($H = 12.975$, 3 d.f., $P < 0.01$). Pairwise multiple comparisons (Dunn's method) revealed significant differences between PV–ChR2 and SOM–ChR2 ($P < 0.05$) and between SOM–ChR2 and SOM–ARCH ($P < 0.05$). In addition, the Wilcoxon signed rank test was used to determine statistical significance of the median CIs for the individual groups and revealed effects for PV–ChR2 ($P < 0.01$) and for SOM–ChR2 ($P < 0.05$).

Extended Data Fig. 1c: comparisons of the J3 and DB values for single-unit cluster quality between optogenetically identified PV⁺ cells ($n = 36$) or SOM⁺ cells ($n = 20$) and random noise clusters in channels without units ($n = 58$), respectively. Owing to a non-normal distribution, the Kruskal–Wallis one-way ANOVA on ranks test was used and revealed significant differences between the different cell groups, both for the J3 ($H = 86.491$, 2 d.f., $P < 0.001$) and the DB ($H = 86.275$, 2 d.f., $P < 0.001$) values. Pairwise multiple comparisons (Dunn's method) revealed significant differences between PV⁺ cells and control and between SOM⁺ cells and control for both J3 and DB (all $P < 0.05$).

Extended Data Fig. 1e: linear correlations were calculated for optogenetically identified PV⁺ cells ($n = 36$), SOM⁺ cells ($n = 20$) and between waveforms from different and random non-light responsive units ($n = 72$) over different behavioural sessions. Plotted are the maximum *r* values.

Extended Data Fig. 2a: comparisons of baseline freezing levels with and without optogenetic stimulation of PV⁺ or SOM⁺ cells. Two-tailed, paired Student's *t*-tests did not reveal any significant differences between baseline freezing and stimulation in any of the conditions ($P = 0.56$ for PV⁺–ChR ($n = 31$), $P = 0.9$ for SOM⁺–ChR ($n = 7$), $P = 0.82$ for PV⁺–ARCH ($n = 17$), $P = 0.27$ for SOM⁺–ARCH ($n = 6$)).

Extended Data Fig. 2b: effects of all performed optogenetic manipulations of either PV⁺ or SOM⁺ cells on the acquisition of fear. For each condition the internal control CS1 was compared with the CS2, which was paired with optogenetic stimulation during conditioning. Two-tailed, paired Student's *t*-tests revealed significant differences for all conditions, except for the CS+US control group ($P < 0.001$ for PV⁺–ChR CS+US ($n = 11$), $P = 0.86$ for PV⁺–ChR CS+US control ($n = 10$), $P < 0.05$ for PV⁺–ChR US ($n = 11$), $P < 0.001$ for PV⁺–ARCH US ($n = 9$), $P < 0.05$ for PV⁺–ChR CS ($n = 10$), $P < 0.05$ for PV⁺–ARCH CS ($n = 8$), $P < 0.01$ for SOM⁺–ChR CS ($n = 10$) and $P < 0.05$ for SOM⁺–ARCH CS ($n = 8$)).

Extended Data Fig. 4d: comparisons of the spike half-width and the spontaneous activity of optogenetically identified PV⁺ cells ($n = 36$), SOM⁺ cells ($n = 20$) and non-light-responsive units ($n = 142$). Owing to a non-normal distribution, a Kruskal–Wallis one-way ANOVA on ranks was used and indicated statistically significant differences between the groups for both parameters (spike half-width: $H = 26.347$, 2 d.f., $P < 0.001$; spontaneous activity: $H = 34.518$, 2 d.f., $P < 0.001$). Pairwise multiple comparisons (Dunn's method) revealed significant differences for both parameters between PV⁺ and non-light responsive cells, as well as between PV⁺ and SOM⁺ cells (all $P < 0.05$).

Extended Data Fig. 6e: comparison of freezing levels to CS1 (control) and CS2 (with optogenetic manipulation during conditioning) for principal neuron inhibition ($n = 8$) during the CS. Two-tailed, paired Student's *t*-tests revealed a significant difference ($P < 0.01$).

Extended Data Fig. 8i: comparison of synaptic contacts made by SOM⁺ and PV⁺ interneurons at principal neuron somata (SOM: $n = 8$; PV: $n = 7$). Two-tailed, paired Student's *t*-test (Welch-corrected) revealed a significant difference ($P < 0.05$).

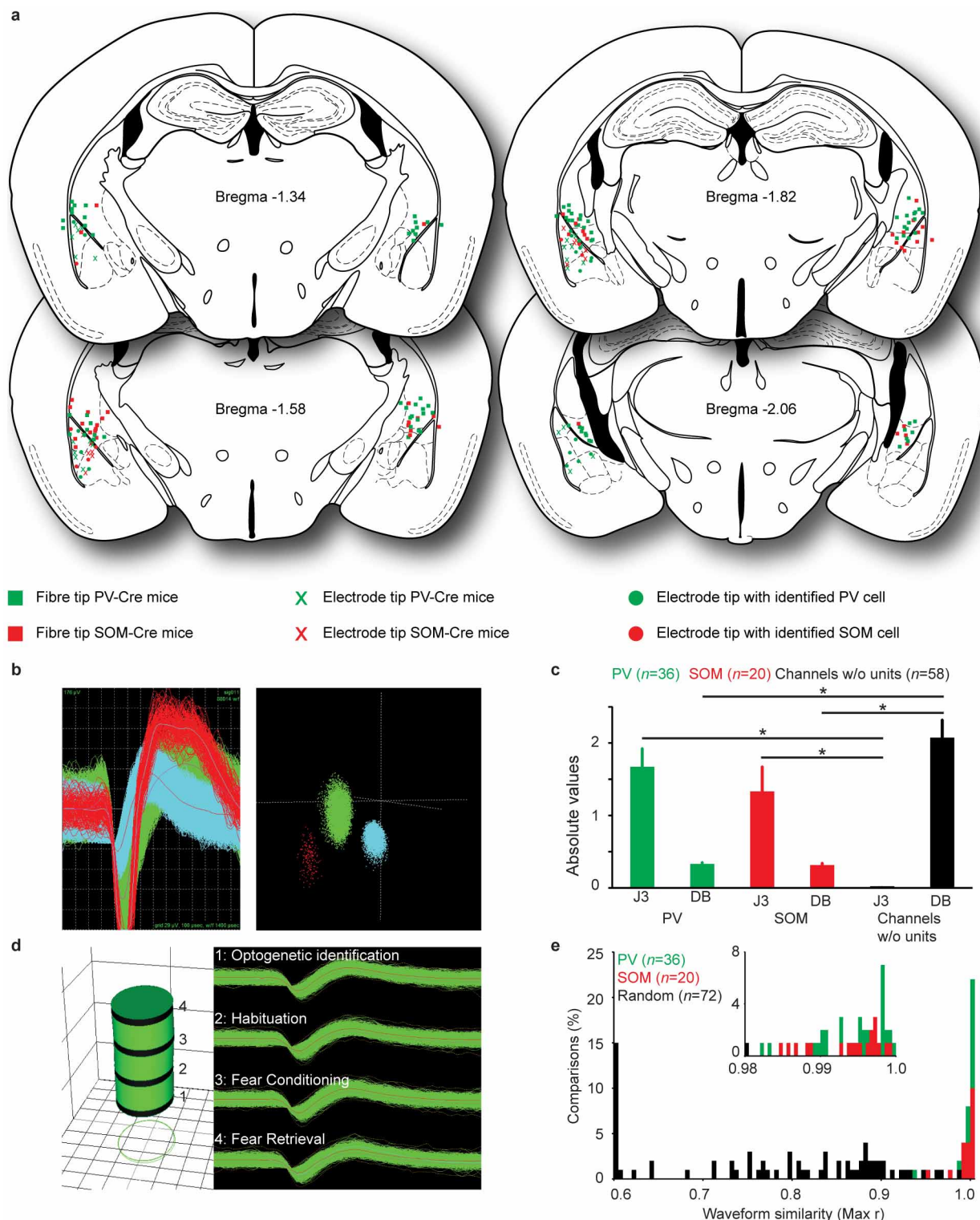
Extended Data Fig. 8j: comparison of extrapolated synaptic contacts made by SOM⁺ and PV⁺ interneurons at the estimated whole dendritic tree of principal neurons (SOM: $n = 8$; PV: $n = 7$). Two-tailed, paired Student's *t*-test revealed a significant difference ($P < 0.001$).

Extended Data Fig. 8k: comparison of synaptic contacts made by SOM⁺ and PV⁺ interneurons at principal neuron dendrites, normalized to the somatic contacts (SOM: $n = 8$; PV: $n = 7$). Two-tailed, paired Student's *t*-test (Welch-corrected) revealed a significant difference ($P < 0.001$).

Extended Data Fig. 8l: comparison of synaptic contacts made by SOM⁺ and PV⁺ interneurons per 50 μ m of principal neuron dendrites, depending on the distance from the soma (SOM: $n = 8$; PV: $n = 7$). Two-tailed, paired Student's *t*-tests (Welch-corrected for the 200 μ m comparison) revealed significant differences (for 50 μ m: $P < 0.001$; for 100 μ m: $P < 0.001$; for 200 μ m: $P < 0.01$).

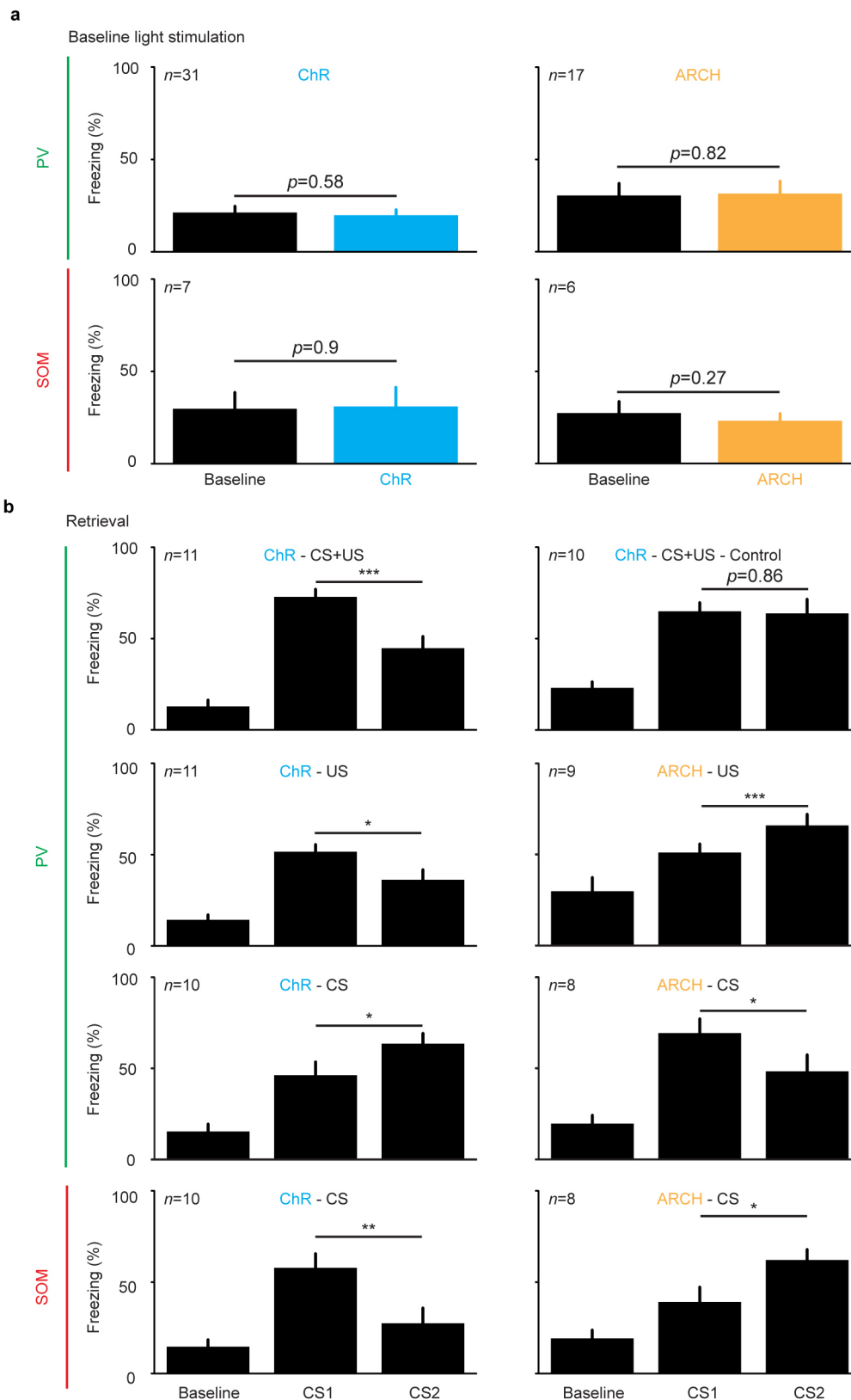
Extended Data Fig. 10b: comparisons of the burst rate and the burst duration in putative principal neurons with and without optogenetic manipulation of PV⁺ or SOM⁺ cells. Bursts were detected using the Poisson surprise method (see above). Data was normalized to control conditions without light. Two-tailed, paired Student's *t*-tests revealed significant differences (for burst rate: $P < 0.05$ for PV–ChR2; $P < 0.001$ for SOM–ChR2; for burst duration: $P < 0.05$ for SOM–ARCH; $P < 0.001$ for PV–ChR2 and SOM–ChR2).

51. Tang, W. *et al.* Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. *J. Neurosci.* **29**, 8621–8629 (2009).
52. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101**, 13124–13131 (2004).
53. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600–606 (2008).
54. Legendy, C. R. & Salzman, M. Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.* **53**, 926–939 (1985).
55. McDonald, A. J. & Mascagni, F. Immunohistochemical characterization of somatostatin containing interneurons in the rat basolateral amygdala. *Brain Res.* **943**, 237–244 (2002).
56. Pivetta, C., Esposito, M. S., Sigrist, M. & Arber, S. Motor-circuit communication matrix from spinal cord to brainstem neurons revealed by developmental origin. *Cell* **156**, 537–548 (2014).
57. Cuntz, H., Forstner, F., Borst, A. & Hausser, M. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Comp. Biol.* **6**, e1000877 (2010).
58. Hill, M. N. *et al.* Disruption of fatty acid amide hydrolase activity prevents the effects of chronic stress on anxiety and amygdalar microstructure. *Mol. Psychiatry* **18**, 1125–1135 (2013).



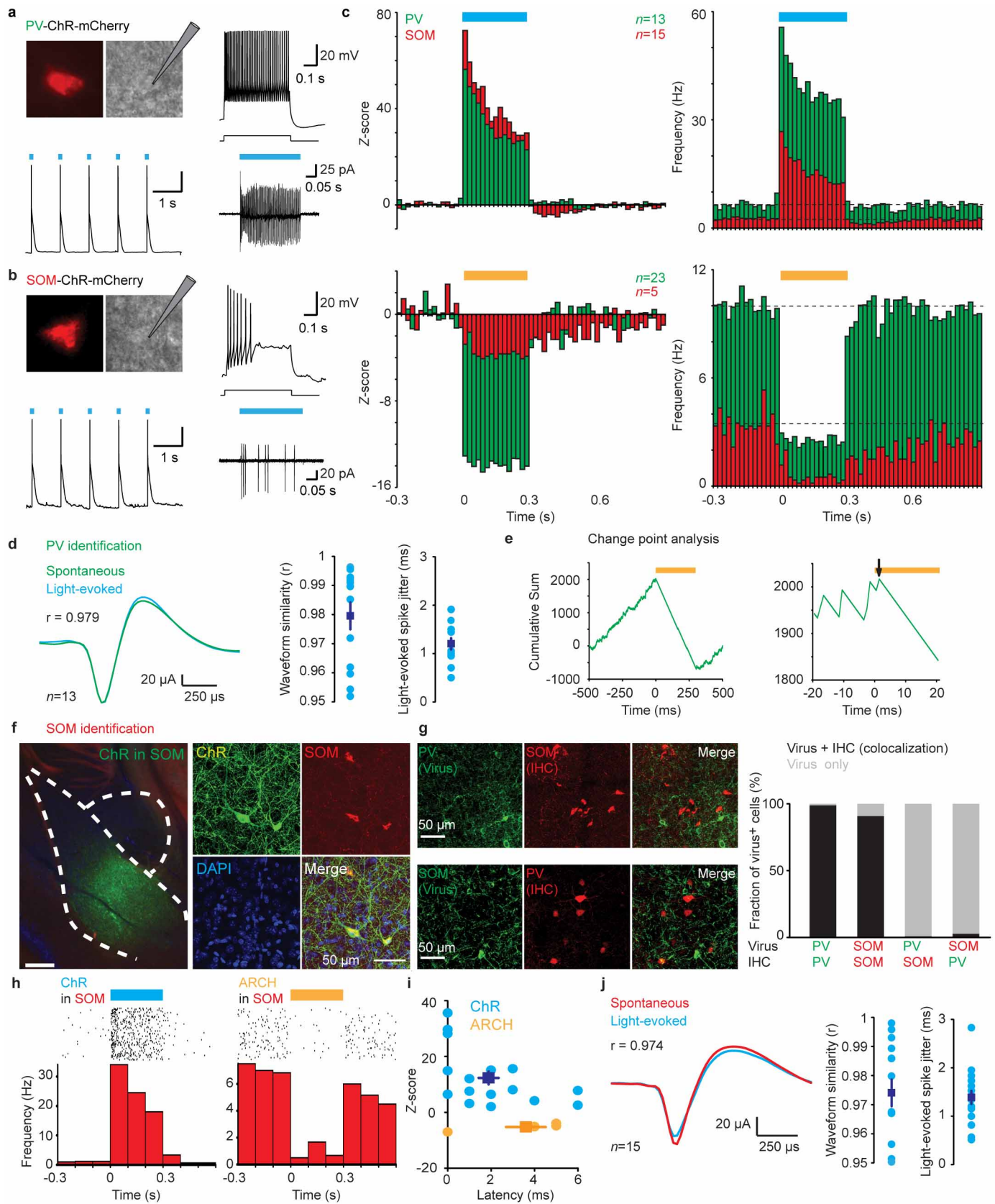
Extended Data Figure 1 | Fibre and electrode placement sites and single-unit recording quality. **a**, Position of electrode tips for single-unit recordings and of bilateral optic fibres for all PV-Cre and SOM-Cre animals used. **b**, Left, superimposed waveforms of three single units recorded on the same channel in the BLA of a PV-Cre animal. Right, spikes from individual units were sorted into clusters, using 3D principal component analysis. **c**, Recording quality was evaluated by calculation of J3 and Davies Bouldin validity index (DB) statistics for PV⁺ and SOM⁺ interneurons. To obtain control values two clusters were selected in the central noise cloud in channels without units. High J3 and low DB values indicate good isolation of single units. **d**, Left, principal component (PC) space cylinders to determine the stability

of clustered waveforms during long-term recordings. Straight cylinders indicate stable recordings of the same set of single units. Right, superimposed waveforms recorded during the different sessions, used for the calculation of the PC space cylinders. **e**, For a quantitative evaluation of spike shape similarity over different recording days, we calculated linear correlation values for PV⁺ and SOM⁺ interneurons. As a control, we calculated the r values for waveforms of different random neurons. The maximum r value across recording sessions was used to quantify similarity ($r = 1$ indicates identical spike shapes). All PV⁺ and SOM⁺ interneurons had r values above 0.95, but only about 4% of the random comparisons. Values are mean \pm s.e.m. * $P < 0.05$. Statistical analysis in Methods.



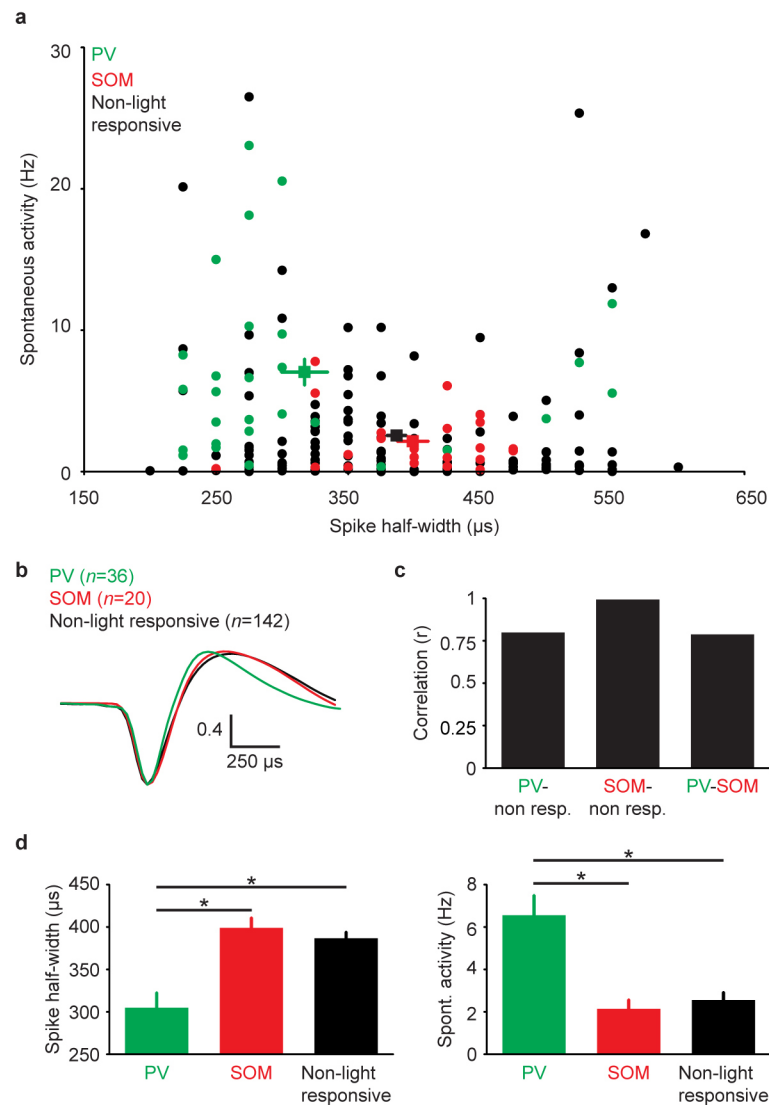
Extended Data Figure 2 | Effects on freezing by baseline optogenetic stimulation and freezing for all behavioural groups. a, Effects of optogenetic manipulations of PV⁺ or SOM⁺ cell activity on baseline freezing levels. Illumination patterns resemble stimulations during the CS in the conditioning paradigms (10 s of 300-ms blue light pulses at 0.9 Hz or 10 s continuous yellow

light). **b,** Freezing values during fear retrieval for all behavioural groups with optogenetic PV⁺ or SOM⁺ cell manipulations during the CS, US or CS+US. Freezing levels are compared between the internal control CS1 and CS2 which was paired with optogenetic stimulation during the conditioning.



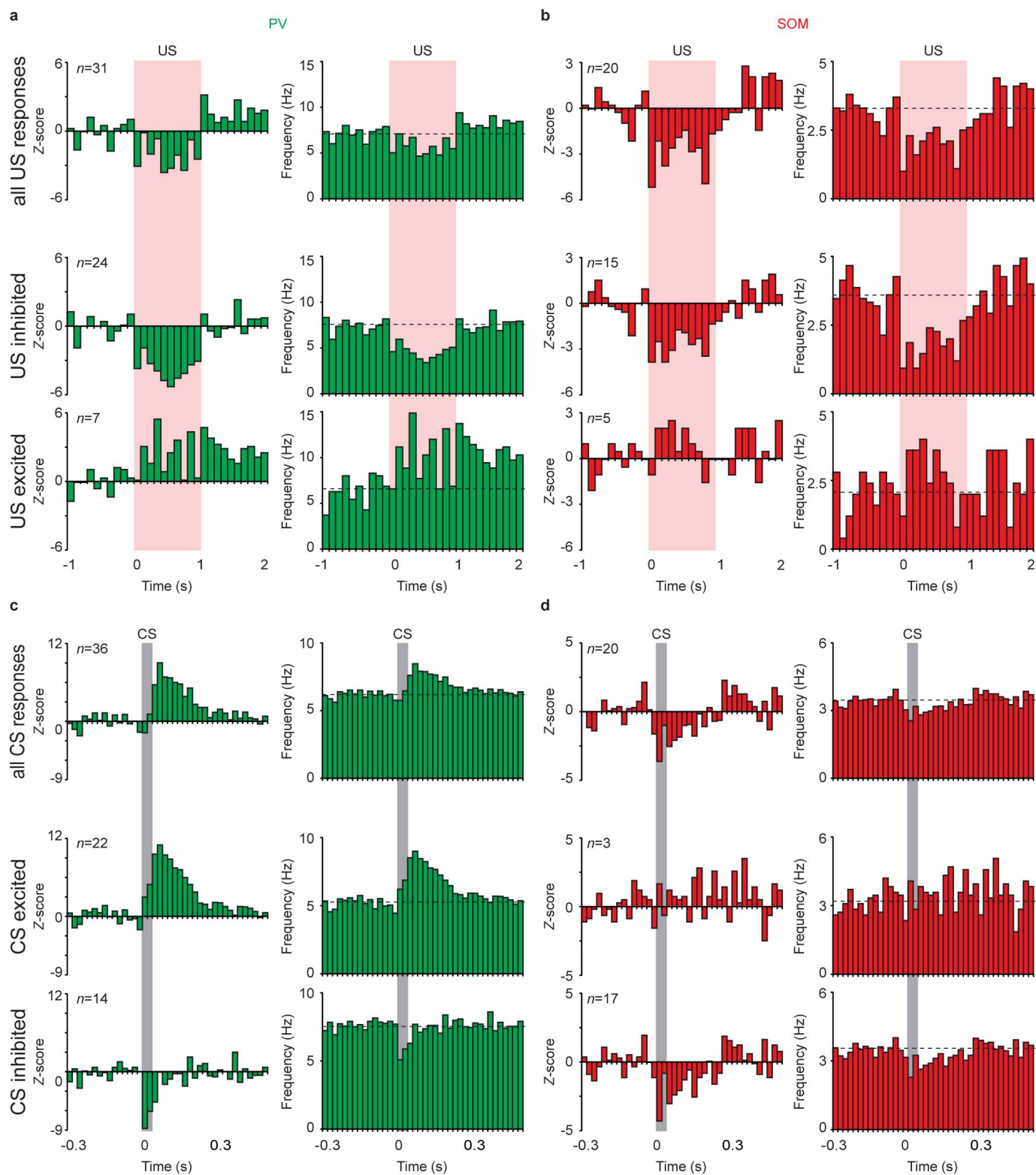
Extended Data Figure 3 | Optogenetic identification and manipulation of PV⁺ and SOM⁺ interneurons. **a, b,** Top left: *ex vivo* whole-cell patch clamp and cell-attached recordings of ChR2-mCherry-expressing PV⁺ (**a**) and SOM⁺ (**b**) BLA interneurons in amygdala slices. Top right: distinct firing patterns of PV⁺ (**a**) and SOM⁺ (**b**) cells in response to depolarizing somatic current injection. Bottom: PV⁺ (**a**) and SOM⁺ (**b**) cells fire brief bursts of action potentials in response to short blue light stimulation (468 nm, 5 ms, 10 mW) and show sustained firing during prolonged stimulation (468 nm, 300 ms, 10 mW, cell-attached recordings). **c,** Single-unit recordings of optogenetically identified PV⁺ (green) and SOM⁺ (red) interneurons upon 300-ms stimulation with blue (ChR2, top) or yellow (ARCH, bottom) light in behaving animals. Left: Z-scored activity. Right: firing frequency. **d,** Left: comparison of spontaneous and light-evoked superimposed average spike waveforms of all PV⁺ interneurons identified by optogenetic activation. Middle: linear correlations between spontaneous and light-evoked spikes were calculated for individual optogenetically identified PV⁺ interneurons. Only cells with *r* values above 0.95 were considered as directly light-activated. Right: jitter of first light-evoked spikes in identified PV⁺ cells. **e,** Change point analysis

for determination of the latency of light-evoked activity changes. The cumulative sum of the activity of a neuron was calculated and the change point was determined. Shown is an example of ARCH-mediated inhibition for the entire illumination period (left) and at the light onset (right). The arrow indicates the change point. **f,** Expression of ChR2, co-expressed with Venus (green), is restricted to SOM⁺ interneurons (red). **g,** Specificity of opsin-expression. Left: examples of opsin expression in a PV-Cre (top) and a SOM-Cre (bottom) animal with immunohistochemistry (IHC) for the respective other interneuron marker. Right: quantification of co-localization of opsin-expressing and IHC-labelled cells. **h,** Light-induced changes in activity of SOM⁺ interneurons upon illumination with either blue or yellow light for activation of ChR2 or ARCH, respectively. **i,** Latencies and Z scores of light-induced activity changes in optogenetically identified SOM⁺ interneurons. **j,** Left: comparison of average waveforms of spontaneous and light-evoked spikes in SOM⁺ interneurons. Middle: linear correlations between spontaneous and light-evoked spikes in optogenetically identified SOM⁺ interneurons. Right: jitter of first light-evoked spikes in identified SOM⁺ cells.



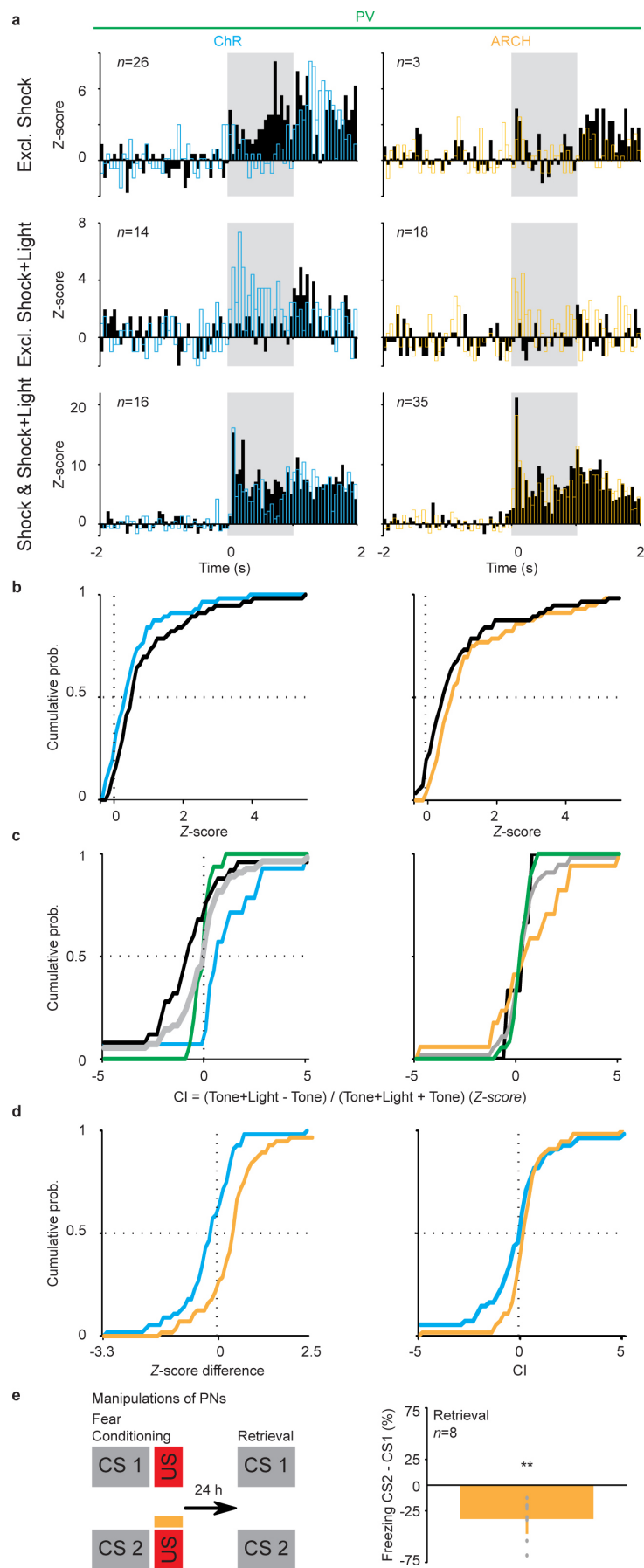
Extended Data Figure 4 | Physiological properties of PV^+ and SOM^+ interneurons and of non-light-responsive cells. **a**, Distribution of physiological properties for optogenetically identified PV^+ and SOM^+ interneurons and for non-light-responsive cells. Spike half-width and spontaneous activity were compared for identified interneurons and for a population of simultaneously recorded non-light-responsive cells, which is mainly comprised of principal cells, but could also include unidentified interneurons. Squares indicate average values for the three groups. SOM^+

interneurons cannot be separated from the non-light-responsive cells whereas PV^+ cells are, on average, different from the other groups. **b**, Comparison of normalized spike waveforms of PV^+ and SOM^+ interneurons and non-light-responsive cells. **c**, Linear correlations of the spike waveforms between the different groups. **d**, Comparison of spike half-width and spontaneous activity of PV^+ and SOM^+ interneurons and non-light-responsive cells. Note the similarity of SOM^+ interneurons with non-light-responsive cells. Values are mean \pm s.e.m. * $P < 0.05$. Statistical analysis in Methods.



Extended Data Figure 5 | US and CS responses in PV⁺ and SOM⁺ interneurons in the BLA. **a–d**, Z-scored population activity (left) and firing frequency (right) for identified PV⁺ (green) and SOM⁺ (red) cells in response to the US (**a**, **b**) or the CS (**c**, **d**). Shown are the population responses for all cells (top graphs). The graphs below show population responses that were

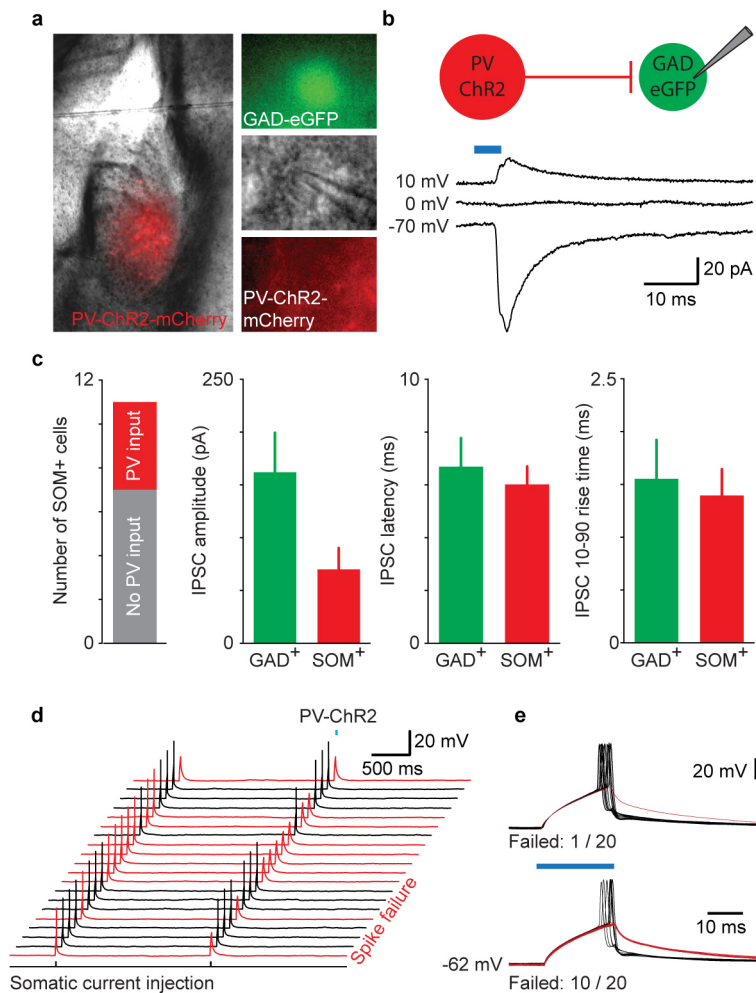
separated according to whether the cells showed inhibitory or excitatory responses to the stimuli. Cells were included in the stimulus inhibited group or stimulus excited group if their average Z-scored stimulus response was <0 or >0, respectively.



Extended Data Figure 6 | Effects of optogenetic manipulations of PV⁺ interneuron activity on footshock responses of putative principal neurons.

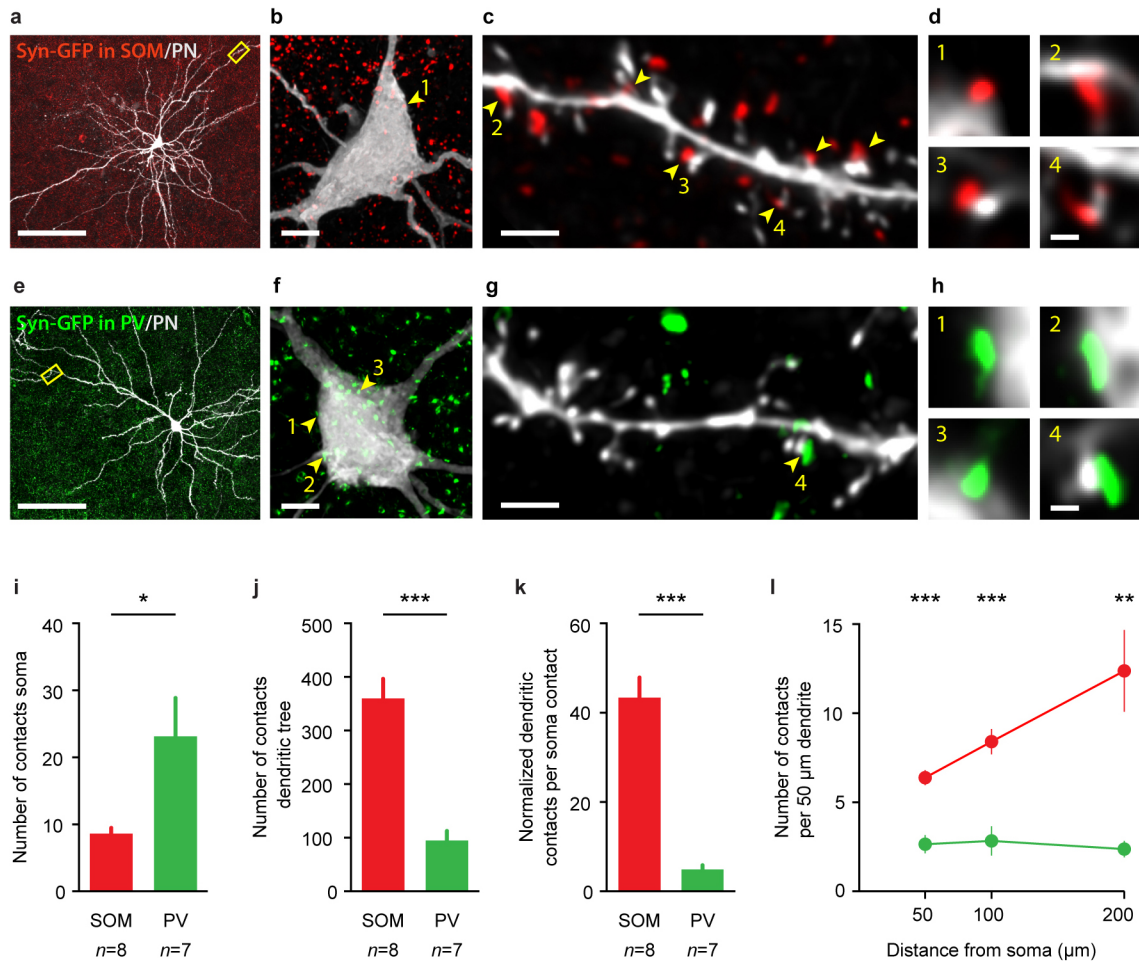
a, Footshock-evoked responses in footshock-excited putative principal neurons with (blue or yellow) and without (black) optogenetic stimulation of PV⁺ interneurons. Cells are split based on being excited either exclusively during the footshock alone, exclusively during the footshock+light or during both conditions. **b**, Cumulative probability distributions of the average Z score of individual putative principal neurons during the footshock without (black) and with optogenetic manipulations of PV⁺ interneurons. **c**, Cumulative probability distributions of the change index (CI) as a measure for the light-induced changes in footshock responses for the different optogenetic interneuron manipulations. The change index is calculated as $CI = (Z \text{ score}(\text{footshock} + \text{light}) - Z \text{ score}(\text{footshock})) / (Z \text{ score}(\text{footshock} + \text{light}) + Z \text{ score}(\text{footshock}))$ (all Z scores are the averages of the shock response during

the entire footshock). Shown are the distributions for the footshock-excited cells (black), for the footshock+light excited cells (blue or yellow), for cells which are excited in both conditions (green) and the weighted sum (grey). **d**, Left, comparisons of cumulative probability distributions for the differences in average footshock Z score during the entire footshock ((footshock+light) – footshock) in putative principal neurons for PV⁺ interneuron activation (blue) and inhibition (yellow). Right, comparisons of cumulative probability distributions for the change index of the footshock responses of putative principal neurons upon optogenetic activation (blue) or inhibition (yellow) of PV⁺ interneurons. **e**, Inhibition of BLA principal neurons during the US impairs fear acquisition. Principal neurons were selectively inhibited by expression of ARCH under the control of the CaMKII promoter and illumination with yellow light during the US. Values are mean \pm s.e.m. ****** $P < 0.01$. Statistical analysis in Methods.



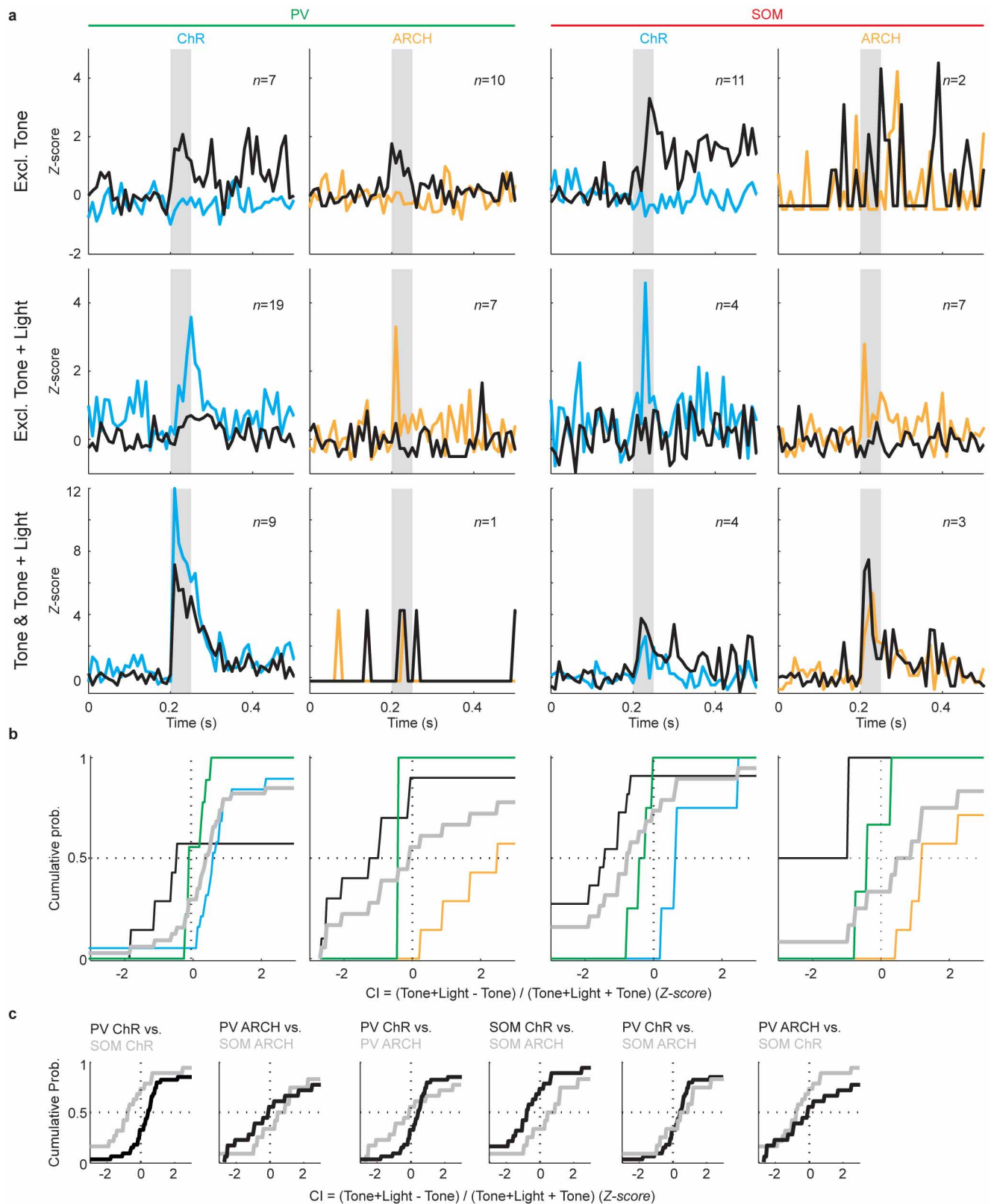
Extended Data Figure 7 | Inhibitory inputs from PV⁺ to SOM⁺ as well as to unidentified GABAergic neurons. **a**, Slice patch-clamp experiments. Left: BLA location of PV⁺ interneurons expressing ChR2-mCherry in acute slices of PV-Cre::GAD-EGFP animals. Right: whole-cell recording (middle) of neuron expressing GAD (top) but not ChR2 (bottom). (Original magnification: left: $\times 5$, right: $\times 60$.) **b**, Optogenetic stimulation of PV⁺ neurons resulted in inhibitory postsynaptic currents (IPSCs) in GAD-EGFP⁺ neurons ($n = 26$, 5-ms blue light). IPSCs reverse at about 0 mV (voltage clamp Cl^- reversal potential, $n = 4$). **c**, Left: fractions of patched and identified SOM⁺ cells with and without detected PV⁺ input. Owing to potential wash-out of the antigen (SOM) during the whole-cell recordings, these data probably represent a solid

lower bound rather than the actual connectivity rate (see Methods). Right: IPSC amplitude, peak latency after blue light onset and 10–90% rise time of PV inputs on all measured GABAergic (PV-GAD, $n = 26$) and somatostatin-positive (PV-SOM, $n = 3$) neurons, respectively. 1 SOM⁺ cell with PV inputs was tested with a current clamp internal at -40 mV and thus does not contribute to the statistics. **d**, **e**, Suppression of spikes in a SOM⁺ cell by optogenetic activation of PV⁺ cells. **d**, individual trials of somatic current injection into a SOM⁺ cell with and without PV⁺ activation. Red traces show spike failures on PV⁺ stimulation. **e**, overlay of current injections with spiking success and failures without (top) and with (bottom) simultaneous PV⁺ cell activation.



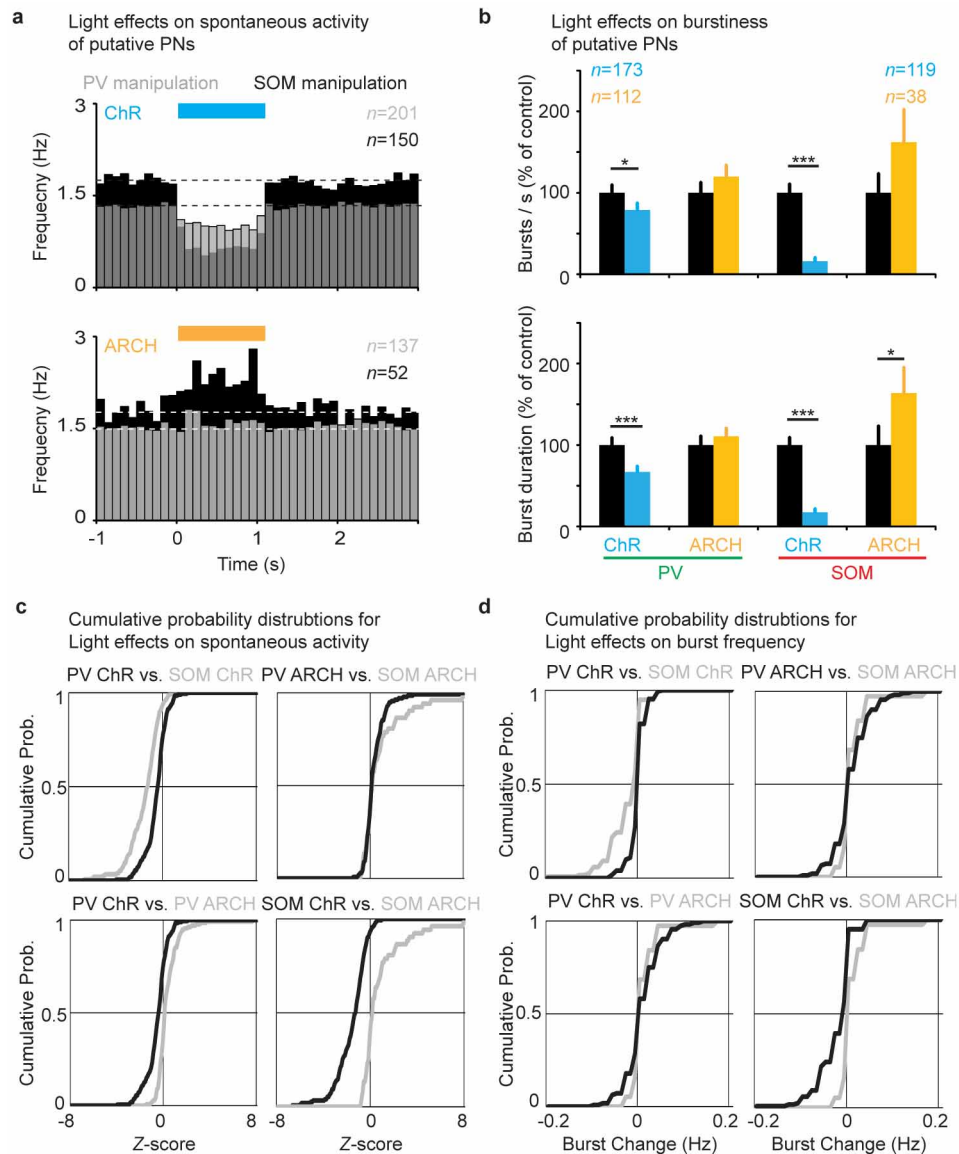
Extended Data Figure 8 | PV^+ and SOM^+ interneurons contact different compartments of pyramidal neurons. **a**, Maximum intensity projection (MIP) showing a biocytin-filled principal neuron (grey) and conditional expression of the presynaptic marker synaptophysin-GFP (Syn-GFP, red) in SOM^+ interneurons in the BLA. **b**, Soma of the same principal neuron (grey) with expression of synaptophysin-GFP (red) selectively in SOM^+ interneurons (MIP). Arrowhead points to 1 out of 7 identified synaptic contacts at the soma revealed by 3D analysis of confocal images. **c**, Dendritic branch located 200 μ m distant from the soma (MIP, yellow rectangle in **a**). Arrowheads mark synaptic contacts of SOM^+ interneurons with the principal neuron dendrite identified in 3D analysis. **d**, Single focal planes of contacts labelled in **b**, **c**. **e**, Biocytin-filled principal neuron and conditional synaptophysin-GFP expression (green) in PV^+ interneurons in the BLA (MIP). **f**, Soma of the same principal neuron with presynaptic synaptophysin-GFP expression selectively in PV^+ interneurons (MIP). Arrowheads point to 3 out of 53 synaptic contacts

from PV^+ interneurons identified in 3D analysis at this soma. **g**, Dendrite located 200 μ m distant from the soma (MIP, yellow rectangle in **e**). Arrowhead marks a synaptic contact from PV^+ interneurons. **h**, Single section images of appositions labelled in **f**, **g**. Scale bars: 100 μ m (**a**, **e**), 5 μ m (**b**, **f**), 2 μ m (**c**, **g**) and 0.5 μ m (**d**, **h**). **i**, Quantification of synaptic contacts from SOM^+ and PV^+ interneurons with the soma of principal neurons in the BLA. **j**, Number of extrapolated synaptic appositions from SOM^+ and PV^+ interneurons onto the estimated dendritic tree of principal neurons. **k**, Normalization of dendritic versus somatic appositions shows a highly significant difference between postsynaptic targeting of SOM^+ and PV^+ interneurons. **l**, Distance dependency of synaptic inputs from SOM^+ (red, $n = 8$) and PV^+ interneurons (green, $n = 7$) onto principal neuron dendrites. SOM^+ interneurons preferentially target distal dendrites of principal neurons. All values are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Statistical analysis in Methods.



Extended Data Figure 9 | Effects of optogenetic manipulations of PV⁺ or SOM⁺ interneuron activity on tone-evoked responses in putative principal neurons. **a**, Tone-evoked responses in tone-excited putative principal neurons with (blue or yellow) and without (black) optogenetic stimulation of PV⁺ or SOM⁺ interneurons. Cells are split based on being tone excited either exclusively during the tone alone, exclusively during tone+light or in both conditions. **b**, Cumulative probability distributions of the change index (CI) as a measure for the light-induced changes in tone responses for the different

optogenetic interneuron manipulations. The change index is calculated as $CI = (Z \text{ score}(\text{tone+light}) - Z \text{ score}(\text{tone})) / (Z \text{ score}(\text{tone+light}) + Z \text{ score}(\text{tone}))$ (all Z scores are the averages of the tone response during the pip). Shown are the distributions for the tone-excited cells (black), for the tone+light excited cells (blue or yellow), for cells which are excited in both conditions (green) and the weighted sum (grey). **c**, Comparisons of cumulative probability distributions for the change index of the tone responses of putative principal neurons upon optogenetic stimulation of PV⁺ or SOM⁺ cells.



Extended Data Figure 10 | Effects of optogenetic manipulations of PV⁺ or SOM⁺ interneurons on spontaneous activity and burst firing of putative principal neurons. **a**, Effects of optogenetic activation (top) or inhibition (bottom) of PV⁺ (grey) or SOM⁺ (black) cells on the spontaneous activity of putative principal neurons. **b**, Effects of optogenetic manipulations of PV⁺ or SOM⁺ cells on the burst frequency (top) and burst duration (bottom) in putative principal neurons. **c**, Comparisons of cumulative probability

distributions for light-induced changes in spontaneous activity in putative principal neurons on optogenetic stimulation of PV⁺ or SOM⁺ interneurons. **d**, Comparisons of cumulative probability distributions for light-induced changes in burst rate in putative principal neurons on optogenetic stimulation of PV⁺ or SOM⁺ interneurons. Values are mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$. Statistical analysis in Methods.

Contrasting forms of cocaine-evoked plasticity control components of relapse

Vincent Pascoli^{1*}, Jean Terrier^{1*}, Julie Espallergues^{2,3,4}, Emmanuel Valjent^{2,3,4}, Eoin Cornelius O'Connor¹ & Christian Lüscher^{1,5}

Nucleus accumbens neurons serve to integrate information from cortical and limbic regions to direct behaviour. Addictive drugs are proposed to hijack this system, enabling drug-associated cues to trigger relapse to drug seeking. However, the connections affected and proof of causality remain to be established. Here we use a mouse model of delayed cue-associated cocaine seeking with *ex vivo* electrophysiology in optogenetically delineated circuits. We find that seeking correlates with rectifying AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor transmission and a reduced AMPA/NMDA (*N*-methyl-D-aspartate) ratio at medial prefrontal cortex (mPFC) to nucleus accumbens shell D1-receptor medium-sized spiny neurons (D1R-MSNs). In contrast, the AMPA/NMDA ratio increases at ventral hippocampus to D1R-MSNs. Optogenetic reversal of cocaine-evoked plasticity at both inputs abolishes seeking, whereas selective reversal at mPFC or ventral hippocampus synapses impairs response discrimination or reduces response vigour during seeking, respectively. Taken together, we describe how information integration in the nucleus accumbens is commandeered by cocaine at discrete synapses to allow relapse. Our approach holds promise for identifying synaptic causalities in other behavioural disorders.

The nucleus accumbens (NAc) is a point of convergence for excitatory afferents arising from limbic and cortical regions, including the basolateral amygdala (BLA), ventral hippocampus (vHipp) and the mPFC. From pharmacology, lesion and imaging experiments, each of these regions is thought to signal distinct information to the NAc during reward-related situations. Specifically, the BLA signals emotional valence, the vHipp provides contextual relevance, whereas the mPFC provides action–outcome information^{1,2}. NAc neurons select and integrate information from these diffuse regions and signal to the basal ganglia motor system to guide appropriate behaviours, such as foraging in response to feeding relevant signals^{3,4}. This neural circuitry is also involved in core features of drug addiction, such as craving and relapse in response to drug-associated cues after withdrawal periods^{5,6}. Understanding how this circuitry is ‘hijacked’ after drug experience and how drug-evoked alterations are causally related to drug-adaptive behaviours, such as relapse, is fundamental to defining the pathophysiology of addiction.

The task of information processing and output in the NAc falls to MSNs, representing 95% of NAc neurons and broadly divisible into two equally sized classes according to the dopamine receptors expressed⁷. Both NAc D1 receptor (D1R)-MSN and D2 receptor (D2R)-MSN subtypes receive excitatory afferents from the BLA, vHipp and mPFC^{4,8}, but differ in their projection targets^{3,9}. Owing to the arrangement of synapses on MSN spines, excitatory transmission arriving onto the spine head is subject to strong modulation by dopamine from the ventral tegmental area arriving at the spine neck¹⁰. Indeed, dopamine gates excitatory transmission, allowing synaptic adaptations and modulation of reward-related behaviours^{11–13}. Accumbal dopamine transients induced by addictive drugs are probably key for inducing circuit adaptations that divert behaviour towards compulsive drug seeking and heighten the risk of relapse after prolonged withdrawal^{14,15}. Relapse is associated with exposure to drug cues, it is context dependent and requires knowledge of what actions result in drug delivery⁵. Therefore, this behaviour is likely underpinned by a memory trace, formed in the reward circuitry

during drug use^{16,17}. However, the nature of this trace and the causal implications remain elusive.

Much literature has described induction requirements and expression mechanisms for drug-evoked plasticity at excitatory synapses, including those of the NAc¹⁷. The time after the last drug exposure is important, as the expression of some forms of plasticity can take days or weeks^{18,19}. Therefore, establishing a causal link between drug-evoked plasticity and drug-adaptive behaviours has been difficult. Pharmacological manipulation of AMPA receptors (AMPA receptors) implicates NAc glutamatergic transmission in cue-associated cocaine seeking^{20,21}, but many questions remain. Do different excitatory inputs change in the same way onto D1R- and D2R-MSNs after withdrawal from cocaine self-administration? Could drug-evoked alterations at specific inputs be causally linked to specific components of drug-adaptive behaviours, such as relapse? Here we use cell-type-specific reporter lines and optogenetics to address these questions, taking an approach that may also be applied in other behavioural disorders to establish synaptic causalities (Extended Data Fig. 1).

Cocaine-evoked plasticity at D1R-MSNs

To characterize drug-evoked plasticity in identified NAc MSNs at a time point when relapse can be observed, bacterial artificial chromosome (BAC) transgenic mice expressing fluorescent proteins under the control of the D1R or D2R promoter were first trained to self-administer intravenous cocaine²². Control mice that received only intravenous saline infusions quickly stopped responding, whereas mice that self-administered cocaine learned to discriminate between the active and inactive levers and responded almost exclusively on the active lever at the end of the acquisition phase, confirming successful acquisition of cocaine self-administration (Fig. 1a, b).

Thirty days after the last self-administration session, whole-cell recordings were made *ex vivo* in identified NAc MSNs with excitatory postsynaptic currents (EPSCs) elicited by electrical stimulation to obtain a global overview of cocaine-evoked plasticity (Fig. 1c). The rectification

¹Department of Basic Neurosciences, Medical Faculty, University of Geneva, CH-1211 Geneva, Switzerland. ²INSERM, U661, Montpellier F-34094, France. ³CNRS, UMR-5203, Institut de Génétique Fonctionnelle, Montpellier F-34094, France. ⁴Universités de Montpellier 1 & 2, UMR-5203, Montpellier F-34094, France. ⁵Clinic of Neurology, Department of Clinical Neurosciences, Geneva University Hospital, CH-1211 Geneva, Switzerland.

*These authors contributed equally to this work.

index of AMPAR-EPSCs was increased selectively in D1R-MSNs from mice that self-administered cocaine, suggesting the presence of GluA2-lacking calcium-permeable AMPARs (CP-AMPA²³). The ratio of the amplitude of AMPAR- and NMDAR-EPSCs (*A/N* ratio) recorded in the same cells, which provides a measure of synaptic strength, was also increased only in D1R-MSNs from mice that self-administered cocaine (Fig. 1c). At first glance, it was surprising to observe these two forms of plasticity in the same neuron, because CP-AMPA²³ show reduced conductance at +40 mV that would produce a decreased *A/N* ratio, if NMDA were unchanged. One potential explanation is that cocaine induced several contrasting forms of plasticity that segregate between different inputs onto the same D1R-MSNs. In support of this idea, a recent study reported an increased *A/N* ratio specifically at vHipp to NAc synapses after chronic non-contingent cocaine injections, without, however, resolving the MSN identity²⁴.

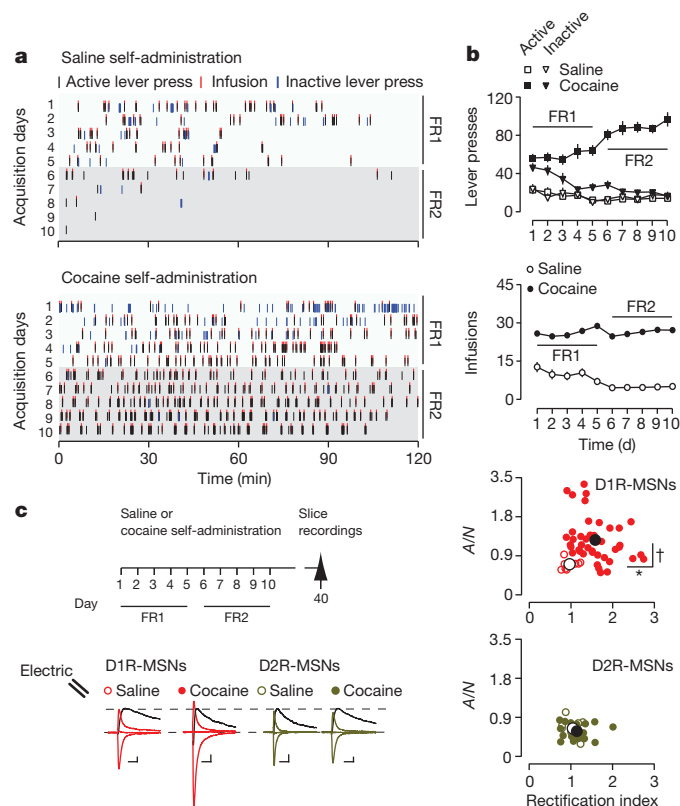


Figure 1 | Withdrawal from cocaine self-administration evokes cell-type-specific modifications of excitatory synapses in the NAc. **a**, Raster plot for infusions, active and inactive lever presses as a function of time during acquisition of self-administration for a mouse that self-administered saline (top) or cocaine (bottom). FR1, FR2, fixed-ratio one and two schedules. **b**, Mean total lever presses (top) and infusions (bottom) during the acquisition phase of saline ($n = 42$) or cocaine ($n = 164$) self-administration for all mice used in the study. **c**, Schematic of experiment (top). Sample traces of AMPAR-EPSCs at -70, 0 and +40 mV (coloured) and NMDAR-EPSCs at +40 mV (black) (bottom) and plots (right) of *A/N* ratio as a function of rectification index (small dots) for each D1R- ($n = 9/41$, saline/cocaine) or D2R-MSN ($n = 8/25$). Large open and filled dots represent group mean saline and cocaine data, respectively. Cocaine increased rectification index and *A/N* ratio in D1R-MSNs but not in D2R-MSNs (rectification index in D1R- and D2R-MSNs: effect of group (cocaine/saline), $F_{1,83} = 10.86$, $P < 0.01$; cell type \times group, $F_{1,83} = 6.08$, $P < 0.05$. Student's *t*-test for cocaine versus saline in D1R-MSNs, $t_{48} = 3.72$, $*P < 0.01$; in D2R-MSNs, $t_{31} = 0.80$, $P = 0.43$. *A/N* ratio in D1R- and D2R-MSNs: effect of group, $F_{1,83} = 5.91$, $P < 0.05$; cell type \times group, $F_{1,83} = 8.34$, $P < 0.01$. Student's *t*-test for cocaine versus saline in D1R-MSNs, $t_{48} = 3.25$, $\dagger P < 0.01$, in D2R-MSNs, $t_{31} = 0.62$, $P = 0.54$). Scale bars, 20 pA, 20 ms. Plots, means with s.e.m.

Cocaine-evoked plasticity at identified inputs

To explore the possibility that distinct afferents to the NAc undergo contrasting forms of plasticity after cocaine self-administration, the optogenetic effector channelrhodopsin tagged with enhanced yellow fluorescent protein (ChR2-eYFP) was virally expressed by stereotaxic injection into the BLA, vHipp or mPFC of BAC transgenic reporter mice (Fig. 2a). Retrograde labelling with cholera toxin subunit B in the dorso- or ventromedial NAc shell confirmed these regions as providing major afferents (Extended Data Fig. 2a), whereas anterograde labelling of different inputs with distinct fluorescent markers identified the juxtaposition of fibres in the NAc (Extended Data Fig. 3), in agreement with studies reporting the convergence of inputs onto the same MSN^{25–27}. Whole-cell recordings from mice infected with ChR2-eYFP in each region confirmed that inputs were excitatory (Extended Data Fig. 2c). Input and cell-type-specific plasticity were then evaluated *ex vivo* in the NAc shell from mice that self-administered cocaine or saline after 1 month of withdrawal.

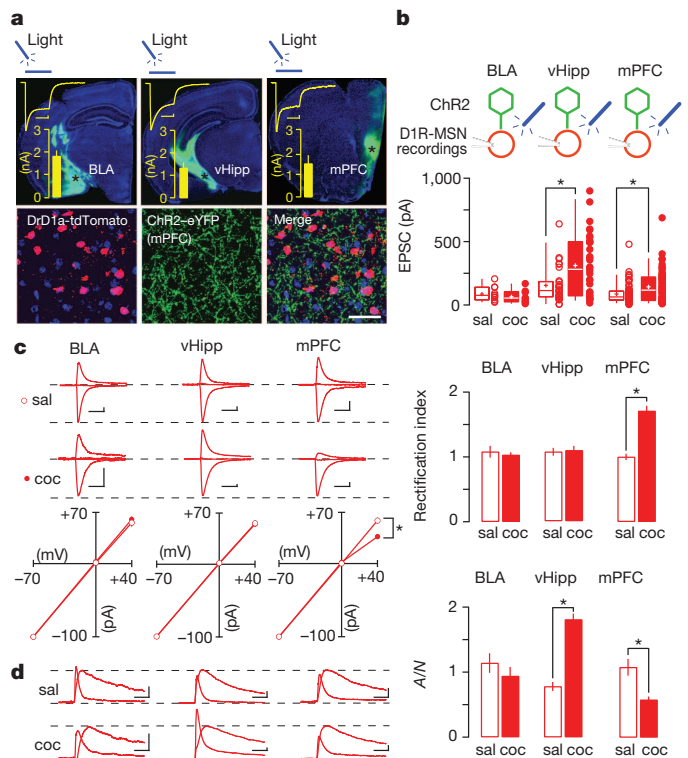


Figure 2 | Withdrawal from cocaine self-administration evokes input-specific plasticity in NAc D1R-MSNs. **a**, Top, images of BLA, vHipp and mPFC infected with ChR2-eYFP, with example traces and mean peak amplitudes of photocurrents (100 ms light pulses). Scale bars, 50 ms, 50 pA. Bottom, confocal image of NAc shell from a Drd1a-tdTomato mouse infected with ChR2 in the mPFC, as used for input and cell-type-specific recordings. Cell nuclei are stained with Hoechst. Scale bar, 100 μ m. **b**, Top, schematic of whole-cell recordings of D1R-MSNs from mice infected with ChR2 in the BLA (left), vHipp (centre) or mPFC (right). Bottom, cocaine (coc) self-administration increased the mean amplitude of 4 ms light-evoked EPSCs at vHipp and mPFC synapses ($t_{62} = 3.12$, $*P < 0.01$, $n = 30/34$ for saline (sal)/cocaine and $t_{120} = 3.0$, $*P < 0.01$, $n = 66/56$, respectively), but did not change BLA synapses ($t_{18} = 0.88$, $P = 0.39$; $n = 11/9$). **c**, For each input, example traces are shown (top), with I/V plots (bottom) and group mean rectification index data (right). Cocaine decreased normalized AMPAR-EPSCs at +40 mV from mPFC ($t_{19} = 4.16$, $*P < 0.01$, $n = 8/13$ for saline/cocaine) but not from BLA or vHipp inputs ($t_{11} = 0.67$, $P = 0.52$, $n = 5/8$ and $t_{20} = 0.18$, $P = 0.86$, $n = 9/13$, respectively). The rectification index was increased at mPFC synapses ($t_{19} = 6.9$, $*P \leq 0.001$), but not at BLA or vHipp synapses ($t_{11} = 0.58$, $P = 0.57$ and $t_{20} = 0.22$, $P = 0.83$, respectively). Scale bars, 20 ms, 20 pA. **d**, The *A/N* ratio was decreased by cocaine self-administration at mPFC synapses ($t_{19} = -3.59$, $P < 0.01$), but increased at vHipp synapses ($t_{20} = 8.33$, $P < 0.001$) and was unchanged at BLA synapses ($t_{11} = 0.94$, $P = 0.36$). $*P < 0.05$. Scale bars, 20 ms, 20 pA. Error bars, s.e.m.

Recordings of light-evoked AMPAR-EPSCs first confirmed that the strongest input onto the NAc arises from the hippocampus²⁴ and that cocaine self-administration evoked a significant potentiation of AMPAR-EPSC amplitudes only at vHipp and mPFC inputs onto D1R-MSNs (Fig. 2b). No change was observed at any input onto D2R-MSNs (Extended Data Fig. 4a). Remarkably, rectification of the current–voltage (*I/V*) curve was detected only at mPFC to D1R-MSN synapses in mice that self-administered cocaine, but not at vHipp or BLA to NAc synapses (Fig. 2c). Thus, cocaine self-administration triggered the insertion of CP-AMPA receptors only at mPFC to D1R-MSN synapses. Confirming findings with electrical stimulation, *I/V* curves were linear for all inputs onto D2R-MSNs in mice that self-administered cocaine (Extended Data Fig. 4b). For the *A/N* ratio, in mice that self-administered cocaine it was decreased at mPFC to D1R-MSN synapses (consistent with the poor conductance of CP-AMPA receptors at +40 mV), but increased at vHipp to D1R-MSN synapses and unaffected at BLA synapses (Fig. 2d). Cocaine self-administration evoked no change in the *A/N* ratio at any input onto D2R-MSNs (Extended Data Fig. 4c). Note that in baseline conditions (that is, mice that self-administered saline) both the rectification index and AMPAR/NMDA ratios were similar across all inputs onto D1R- and D2R-MSNs, as predicted from electrical recordings. Collectively, these results demonstrate that cocaine self-administration and withdrawal results in contrasting forms of plasticity at specific inputs onto the same accumbal D1R-MSN.

Homosynaptic effects of optogenetic protocols

Having identified specific inputs onto NAc shell D1R-MSNs that express drug-evoked plasticity, we wanted to test for a causal link to cue-associated

drug seeking. To this end, we aimed to develop optogenetic protocols *ex vivo* that could restore basal synaptic transmission (Fig. 3a, f). However, more than a complete erasure of cocaine-evoked plasticity, we sought to parse the relationship between plasticity at identified synapses to cue-associated seeking behaviour.

Because synapses were essentially potentiated after cocaine self-administration, optogenetic protocols were used that could trigger long-term depression (LTD). When recording light-evoked AMPAR-EPSCs selectively at vHipp to D1R-MSNs, application of a 1 Hz NMDAR-dependent LTD protocol²⁸ significantly depressed EPSCs both in saline and in cocaine self-administration groups (Fig. 3b). At the same input, a 13 Hz extrasynaptic metabotropic glutamate receptor (mGluR)-dependent LTD protocol (Extended Data Fig. 5a²⁹) triggered a depression in the cocaine self-administration group that was significantly attenuated compared with controls (Fig. 3c). The consequence of applying either protocol on cocaine-evoked plasticity at vHipp to NAc synapses (that is, homosynaptic effects) was then examined. As described above, CP-AMPA receptors were not present at vHipp to NAc synapses after cocaine self-administration and neither the 1 nor 13 Hz vHipp protocol further altered the linear *I/V* curve or rectification index at this input (Fig. 3d). Remarkably, however, the 1 Hz, but not the 13 Hz, vHipp protocol normalized the cocaine-evoked increase in the *A/N* ratio at this input (Fig. 3e). Note that in the group that self-administered saline, the 1 or 13 Hz vHipp protocols had little impact on the *A/N* ratio at these synapses, which was already low at baseline (Fig. 3e).

These experiments were repeated, but now isolating mPFC to D1R-MSN synapses (Fig. 3f). The 1 Hz mPFC protocol only slightly depressed AMPAR-EPSCs in the cocaine self-administration group, whereas the

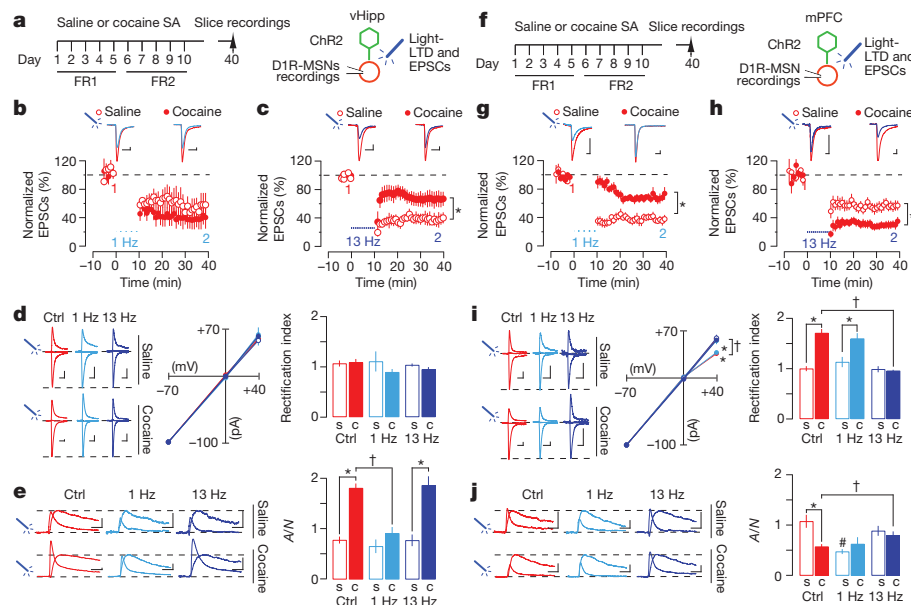


Figure 3 | Homosynaptic effects of optogenetic protocols applied *ex vivo* on D1R-MSN plasticity. **a**, Schematic of experiment for **b–e**; ChR2 in vHipp. **b**, Graph of normalized light-evoked EPSCs across time (bottom: each point represents mean of six sweeps), with example traces (top: mean of 20 sweeps) before (1) and after (2) the LTD protocol (4 ms pulses at 1 Hz, 10 min). One month after cocaine self-administration, the efficiency of the 1 Hz LTD protocol was not modified ($56.3 \pm 14.7\%$ to $38.6 \pm 12.2\%$, $t_9 = 0.89$, $P = 0.4$; $n = 5/4$). **c**, As for **b**, but with a 13 Hz protocol (4 ms pulses at 13 Hz, 10 min). The efficiency of this protocol was reduced in mice that self-administered cocaine ($39.2 \pm 6.7\%$ to $69.5 \pm 10.1\%$, $t_9 = 2.57$, $*P = 0.03$; $n = 6/5$). **d**, Example traces (left) of light-evoked AMPAR-EPSCs recorded at -70 , 0 and $+40$ mV, with *I/V* plot (middle) and group mean rectification index data (right) before (Ctrl) and after the 1 or 13 Hz light protocol, both in saline (s) and cocaine (c) self-administered groups. AMPAR composition was not modified by cocaine or light protocols (saline/cocaine for Ctrl ($n = 9/13$), 1 Hz (6/4), 13 Hz (8/5)). **e**, For the same cells in **d**, example traces (left), with mean *A/N*

ratios (right) before or after the 1 or 13 Hz protocol. Planned comparisons, after analysis of variance (ANOVA), by *t*-test, $\dagger, *P \leq 0.05$. **f**, Schematic of experiment for **g–j**; ChR2 in the mPFC. **g**, The 1 Hz protocol efficiency was reduced in mice that self-administered cocaine ($43 \pm 5.5\%$ to $74 \pm 3.7\%$, $t_{17} = 4.9$, $*P < 0.001$; $n = 8/11$). **h**, The 13 Hz protocol efficiency was increased in mice that self-administered cocaine ($54 \pm 6.1\%$ to $29 \pm 4.7\%$, $t_{22} = 3.25$, $*P < 0.01$; $n = 13/11$). **i**, The 13 Hz but not 1 Hz protocol normalized cocaine-evoked changes in AMPAR-EPSCs at $+40$ mV and rectification index. Planned comparisons, after ANOVA, by *t*-test, $\dagger, *P \leq 0.05$; (saline/cocaine for Ctrl ($n = 9/13$), 1 Hz (10/7), 13 Hz (10/6)). **j**, For the same cells as **i**, the cocaine-evoked decrease in *A/N* ratio was normalized by the 13 Hz, but not the 1 Hz, protocol. The 1 Hz protocol decreased *A/N* ratio in saline mice ($\#P < 0.05$). Planned comparisons, after ANOVA, by *t*-test, $\dagger, *P \leq 0.05$. Note that control data are the same as shown in Fig. 2. Scale bars, 20 pA, 20 ms. Error bars, s.e.m.

depression was significantly larger in controls (Fig. 3g). This finding concurs with a previous study where NMDAR-dependent LTD was altered in 'cocaine-addicted' rats³⁰. The 13 Hz mPFC protocol was actually more efficient in the cocaine self-administration group (Fig. 3h), which is of particular interest because mGluR activation can remove CP-AMPA receptors at many synapses throughout the brain^{31–33}. Both protocols triggered comparable LTD of AMPAR-EPSCs in D2R-MSNs from saline and mice that self-administered cocaine (Extended Data Fig. 5b). The homosynaptic effects of these protocols on cocaine-evoked plasticity at mPFC to NAc synapses was then examined which, as described above, indicated the presence of CP-AMPA receptors. Consistent with LTD experiments, the 13 Hz, but not the 1 Hz, mPFC protocol normalized both the rectifying *I/V* curve and the depressed *A/N* ratio (Fig. 3i, j), pointing to removal of CP-AMPA receptors from mPFC to D1R-MSN synapses. In the group that self-administered saline, the 1 Hz, but not the 13 Hz, mPFC protocol significantly reduced the *A/N* ratio, whereas the rectification index remained unchanged in either case.

Heterosynaptic effects of optogenetic protocols

After evaluating the 1 and 13 Hz protocols on light-evoked homosynaptic transmission, the possibility of heterosynaptic effects was explored *ex vivo* (that is, normalization of cocaine-evoked plasticity at an input other than which the protocol was applied). Indeed, *in vivo* recordings have demonstrated that activation of one excitatory NAc input can trigger heterosynaptic plasticity²⁷.

Optogenetic LTD protocols were applied onto slices at vHipp or mPFC to NAc inputs but EPSCs were evoked with electrical stimulation, thus recruiting transmission from multiple afferents (Fig. 4a). The cocaine-evoked rectification of AMPAR transmission was restored by the 13 Hz protocol, whether applied at vHipp or mPFC to NAc synapses (Fig. 4b). Because rectification was only found at mPFC to D1R-MSN synapses, the 13 Hz vHipp protocol must have triggered normalization of this synapse heterosynaptically, most probably by activating mGluRs

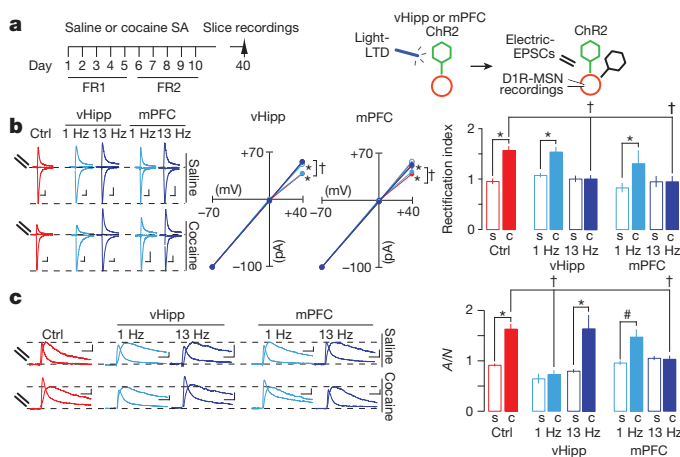


Figure 4 | Heterosynaptic effects of optogenetic protocols applied *ex vivo* on D1R-MSN plasticity. **a**, Schematic of experiment. **b**, Example traces (left) of electrically evoked AMPAR-EPSCs recorded at -70 , 0 and $+40$ mV, with *I/V* plots (middle) and grouped mean rectification index data (right) before (Ctrl) or after light protocols (1 or 13 Hz) were applied with ChR2 in the vHipp or mPFC. The 13 Hz but not 1 Hz light protocol applied at vHipp or mPFC synapses normalized AMPA-EPSCs at $+40$ mV and rectification index (planned comparisons, after ANOVAs, with *t*-tests: $^*P \leq 0.05$; (saline/cocaine for Ctrl ($n = 9/41$), for vHipp 1 Hz (6/8) and 13 Hz (8/8), for mPFC 1 Hz (12/9) and 13 Hz (15/8))). **c**, For the same cells shown in **b**, example traces (left) and group mean *A/N* ratios (right) before or after the 1 or 13 Hz protocols. The 1 Hz vHipp or the 13 Hz mPFC protocol normalized the cocaine-evoked increase in *A/N* ratio, whereas the *A/N* ratio remained elevated after the 13 Hz vHipp protocol or the 1 Hz mPFC protocol (planned comparisons, after ANOVAs, with *t*-tests: $^*P < 0.05$ and $^{\dagger}P = 0.07$). Note that control data are the same as shown in Fig. 1. Scale bars, 20 pA, 20 ms. Error bars, s.e.m.

at neighbouring synapses through glutamate spillover or intracellular signalling. For the *A/N* ratio, which was increased in the cocaine self-administration group when recorded with electrical stimulation, normalization was observed using the 1 Hz vHipp protocol (Fig. 4c). This probably reflected a homosynaptic effect as reported above (see Fig. 3e). However, the 13 Hz mPFC protocol, in addition to removing rectifying AMPARs, also restored the *A/N* ratio; a heterosynaptic effect probably occurring through glutamate spillover activating NMDARs at vHipp to D1R-MSN synapses. Note that the 13 Hz vHipp protocol, in contrast to its heterosynaptic effect on the rectification index, did not elicit a homosynaptic reduction in the *A/N* ratio, possibly because NMDAR activation was too strong to elicit depression^{34,35}. Because our initial validations were performed *ex vivo*, we also confirmed that optogenetic protocols were efficient in normalizing cocaine-evoked plasticity when applied *in vivo* (Extended Data Figs 6 and 7, which include extra discussion). Taken together, this approach led to an advantageous situation to test the causality of contrasting forms of drug-evoked plasticity in cue-associated cocaine seeking, because plasticity at mPFC or vHipp to D1R-MSNs could be restored separately using the 13 Hz vHipp or the 1 Hz vHipp protocol, respectively, or together by using the 13 Hz mPFC protocol.

Causal link to components of cocaine seeking

Now we applied optogenetic protocols in self-administration mice to test directly the causality between drug-evoked plasticity at identified inputs and a drug-adaptive behaviour relevant to addiction. After 1 month of withdrawal from self-administration, mice underwent a cue-associated seeking test, a rodent model of relapse, with optogenetic protocols applied once, 4 h before the test (Fig. 5a, b).

The 1 Hz vHipp protocol (that is, reversing cocaine-evoked plasticity only at vHipp to D1R-MSNs) significantly reduced cue-associated lever pressing, a measure of response vigour, although cocaine seeking per se was still present. The 13 Hz vHipp protocol (that is, inducing a heterosynaptic reversal of plasticity at mPFC to D1R-MSNs) led to increased inactive lever responding, pointing to impaired action–outcome response

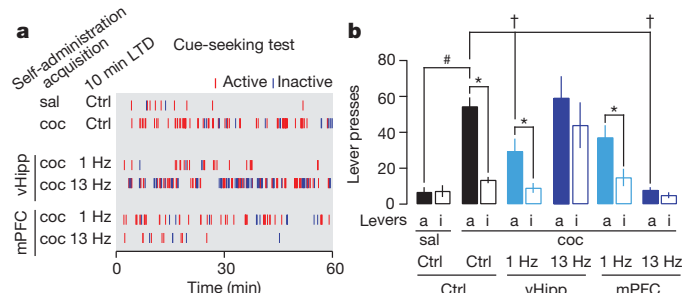


Figure 5 | Linking cocaine-evoked plasticity at identified inputs to specific components of cue-associated cocaine seeking. **a**, Raster plots showing active and inactive lever presses during a cue-associated seeking session. **b**, Group mean data showing active (a) and inactive (i) lever presses. Seeking was robust in control mice that self-administered cocaine (active versus inactive lever, $t_{43} = 9.5$, $^*P < 0.001$; $n = 11/44$, saline/cocaine; active lever to saline control group, after ANOVA, by *t*-test: $t_{53} = 4.6$, $^{\dagger}P < 0.001$). Seeking was present after the 1 Hz vHipp protocol ($t_{10} = 3.6$, $^*P < 0.01$), but active lever responses were significantly diminished (versus cocaine control group, after ANOVA, by *t*-test: $t_{53} = 2.3$, $^{\dagger}P < 0.05$; $n = 11/44$). Lever discrimination was lost after the 13 Hz vHipp protocol ($t_8 = 1.99$, $P > 0.05$), and inactive lever responses were significantly increased (versus cocaine control group, after ANOVA, by *t*-test: $t_{51} = 4.48$, $P < 0.001$; $n = 9/44$). Seeking was present after the 1 Hz mPFC protocol and did not differ from the cocaine controls (versus cocaine control groups by ANOVA: lever, protocol and lever \times protocol interaction all non-significant; $n = 11/44$). Active lever responding was reduced after the 13 Hz mPFC protocol (versus cocaine control group, after ANOVA, by *t*-test: $t_{52} = 4.3$, $^*P < 0.001$; $n = 10/44$) and lever discrimination was lost (active versus inactive levers by *t*-test, $t_9 = 1.7$, $P > 0.1$, $n = 10$). See Extended Data Table 1 for complete statistics. Error bars, s.e.m.

discrimination. The 1 Hz mPFC protocol (that is, no effect on cocaine-evoked plasticity) was without effect on seeking behaviour. Finally, the 13 Hz mPFC protocol (that is, normalizing both inputs) abolished seeking behaviour (Fig. 5a, b). To examine the persistence of this effect, the same mice were tested for cue-associated cocaine seeking 1 week later without further optogenetic intervention. Seeking behaviour remained absent in mice that received the 13 Hz mPFC protocol (Extended Data Figs 8a, b and 9). Taken together, these data demonstrate that cocaine-evoked plasticity at two inputs onto NAc D1R-MSNs is required for the complete expression of cue-associated seeking, whereas each form of contrasting plasticity is causally related to a different component of the seeking behaviour, namely the vigour of the seeking response and the ability to discriminate cocaine- from non-cocaine-directed actions.

We performed extra tests to control for the behavioural specificity of optogenetic interventions. First, optogenetic removal of cocaine-evoked plasticity did not preclude the acquisition of further reward-related instrumental learning (Extended Data Fig. 8c, which includes further description). Second, the 13 Hz mPFC protocol did not affect cue-associated food seeking behaviour in a cohort of mice that were previously trained to self-administer sucrose pellets (Extended Data Fig. 8d–f, which includes further description), suggesting that this protocol is selective for cocaine-evoked plasticity necessary for cocaine seeking against plasticity that may correlate with food seeking^{36,37}.

Discussion

A striking feature of this study is the input-specific expression mechanism of drug-evoked plasticity. In the same D1R-MSNs, synapses that belong to the mPFC input show rectifying AMPAR-EPSCs reflecting the insertion of GluA2-lacking CP-AMPA receptors, whereas at neighbouring vHipp synapses an increased A/N ratio and significantly larger EPSC amplitude indicates the insertion of more GluA2-containing AMPARs. Moreover, such forms of plasticity were absent at other synapses and in D2R-MSNs, although these neurons may still contribute to features of addiction³⁸. Why are some synapses susceptible to cocaine-evoked plasticity and why do the expression mechanisms differ between inputs? These findings may reflect a complex induction process, requiring specific neuronal activity patterns coinciding with high levels of mesolimbic dopamine during drug exposure. Indeed, different drug self-administration histories may favour the induction of synaptic plasticity at other NAc inputs, such as the BLA³⁹. Another possibility is that contrasting expression mechanisms reflect differences in the quality of basal synaptic transmission between cortical and limbic inputs onto NAc MSNs²⁴, or differences in the sensitivity of synapses to be influenced by the coincidence of converging glutamate and dopamine signals involving D1R signalling^{11,27}.

Here we focus on the vHipp and mPFC input onto D1R-MSNs, taking advantage of endogenous depression mechanisms (for example, NMDAR or mGluR activation^{28,29,40}) to restore normal transmission in each input separately or both inputs together. Although the 1 Hz LTD protocol is sufficient to recruit NMDARs on the activated input, trains of action potentials (for example, 13 Hz stimulation) are required to release sufficient amounts of glutamate to activate perisynaptically located mGluRs, even at neighbouring inputs. mGluR-LTD in the NAc is expressed both pre- and postsynaptically^{29,33,41}, but what most likely matters is that mGluR-LTD is an efficient mechanism to remove GluA2-lacking AMPARs, just as in other central nervous system synapses^{31,32,42}. It is therefore not surprising that the 1 Hz mPFC protocol failed to normalize rectification index and had no behavioural effect. The 1 Hz vHipp protocol normalized the A/N ratio at vHipp to D1R-MSN synapses and significantly reduced the vigour with which the animal pressed the active lever in search of cocaine. Thus, cocaine-evoked plasticity at vHipp afferents to NAc D1R-MSNs may enhance recognition of the correct context in which cocaine can be obtained. In the absence of this plasticity, context recognition may be impaired and the certainty of actions reduced. Meanwhile, removal of CP-AMPA receptors from mPFC to D1R-MSN synapses led to a failure of discrimination between the cocaine-associated lever and a second lever on which responding had no consequence. This

establishes cocaine-evoked plasticity at mPFC afferents to NAc D1R-MSNs as necessary to allow correct action–outcome selection during seeking. Finally, restoring transmission in both inputs abolished cue-induced cocaine seeking. Thus, cocaine-evoked plasticity at multiple inputs, with distinct expression mechanisms that change both the efficacy and quality of transmission, are collectively necessary for the full expression of cue-associated drug-seeking behaviour.

Previous studies have recognized the importance of NAc excitatory transmission in drug-adaptive behaviours using systemic or local applications of pharmacological agents that are active during the behavioural test^{20,21,43}. However, a key difference in the present study is that optogenetic protocols were applied outside the test and remained effective 1 week later. Thus, findings here provide support for a model whereby a cocaine memory trace commanders accumbal integration to control core components of relapse. Co-opting endogenous plasticity mechanisms emerges as a radically new way to modify behaviour and may serve as a blueprint for defining synaptic causalities in other synaptic disorders.

METHODS SUMMARY

Mice learned to self-administer cocaine or saline during once-daily 2 h sessions for 10 consecutive days²². In each self-administration session, mice could respond on either an active lever that resulted in a single infusion (0.75 mg kg^{−1} cocaine or saline) paired with a 5 s cue light and followed by a 20 s time-out period, or a second inactive lever that had no consequence. Responding was maintained under a fixed-ratio one schedule (one active lever press gives one infusion) during sessions 1–5, and fixed-ratio two schedule during sessions 6–10. After the final self-administration session, mice were retained in their home cage for 1 month of forced withdrawal. The 60 min test of cue-associated cocaine seeking took place in the same box where self-administration training occurred. Responses on the active lever now triggered a 5 s illumination of the cocaine-associated cue light, but without intravenous infusion, whereas inactive lever responses had no consequence. Optogenetic protocols were applied once *in vivo*, in the home cage, through bilaterally implanted optical fibres targeting the NAc shell 4 h before the seeking test, or before animals were killed for *ex vivo* recordings. For whole-cell patch clamp recordings, the rectification index was calculated after measuring light-evoked EPSC amplitudes at different holding potentials (−70, 0 and +40 mV) in the presence of (D)-2-amino-5-phosphonovaleric acid (AP5; 50 μM). The A/N ratio was calculated at +40 mV with the AMPAR component pharmacologically isolated using AP5 (50 μM) and the NMDAR-EPSC component determined by subtraction. When using ChR2–eYFP in BAC transgenic mice, D2R-MSNs were identified as D2R–eGFP (enhanced green fluorescent protein)-positive neurons in Drd2–eGFP mice or as tomato-negative neurons in Drd1a–tdTomato mice, whereas D1R-MSNs were always identified as tomato-positive cells. This approach was validated in a previous report²⁸.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 17 March 2014.

- Robbins, T. W. & Everitt, B. J. Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* **6**, 228–236 (1996).
- Berridge, K. C. & Kringelbach, M. L. Neuroscience of affect: brain mechanisms of pleasure and displeasure. *Curr. Opin. Neurobiol.* **23**, 294–303 (2013).
- Humphries, M. D. & Prescott, T. J. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog. Neurobiol.* **90**, 385–417 (2010).
- Papp, E. *et al.* Glutamatergic input from specific sources influences the nucleus accumbens-ventral pallidum information flow. *Brain Struct. Funct.* **217**, 37–48 (2012).
- Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neurosci.* **8**, 1481–1489 (2005).
- Kalivas, P. W. & Volkow, N. D. The neural basis of addiction: a pathology of motivation and choice. *Am. J. Psychiatry* **162**, 1403–1413 (2005).
- Gangarossa, G. *et al.* Distribution and compartmental organization of GABAergic medium-sized spiny neurons in the mouse nucleus accumbens. *Front. Neural. Circuits* **7**, 22 (2013).
- MacAskill, A. F., Little, J. P., Cassel, J. M. & Carter, A. G. Subcellular connectivity underlies pathway-specific signaling in the nucleus accumbens. *Nature Neurosci.* **15**, 1624–1626 (2012).
- Smith, R. J., Lobo, M. K., Spencer, S. & Kalivas, P. W. Cocaine-induced adaptations in D1 and D2 accumbens projection neurons (a dichotomy not necessarily synonymous with direct and indirect pathways). *Curr. Opin. Neurobiol.* **23**, 546–552 (2013).

10. Freund, T. F., Powell, J. F. & Smith, A. D. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience* **13**, 1189–1215 (1984).
11. Goto, Y. & Grace, A. A. Limbic and cortical information processing in the nucleus accumbens. *Trends Neurosci.* **31**, 552–558 (2008).
12. Cerovic, M., d'Isa, R., Tonini, R. & Brambilla, R. Molecular and cellular mechanisms of dopamine-mediated behavioral plasticity in the striatum. *Neurobiol. Learn. Mem.* **105**, 63–80 (2013).
13. Saddoris, M. P., Sugam, J. A., Cacciapaglia, F. & Carelli, R. M. Rapid dopamine dynamics in the accumbens core and shell: learning and action. *Front. Biosci. (Elite Ed.)* **5**, 273–288 (2013).
14. Schultz, W. Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. *Neuron* **69**, 603–617 (2011).
15. Di Chiara, G. & Bassareo, V. Reward system and addiction: what dopamine does and doesn't do. *Curr. Opin. Pharmacol.* **7**, 69–76 (2007).
16. Hyman, S. Addiction: a disease of learning and memory. *Am. J. Psychol.* **162**, 1414–1422 (2005).
17. Lüscher, C. & Malenka, R. C. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. *Neuron* **69**, 650–663 (2011).
18. Kourrich, S., Rothwell, P. E., Klug, J. R. & Thomas, M. J. Cocaine experience controls bidirectional synaptic plasticity in the nucleus accumbens. *J. Neurosci.* **27**, 7921–7928 (2007).
19. Ortinski, P. I., Vassoler, F. M., Carlson, G. C. & Pierce, R. C. Temporally dependent changes in cocaine-induced synaptic plasticity in the nucleus accumbens shell are reversed by D1-like dopamine receptor stimulation. *Neuropsychopharmacology* **37**, 1671–1682 (2012).
20. Cornish, J. L. & Kalivas, P. W. Glutamate transmission in the nucleus accumbens mediates relapse in cocaine addiction. *J. Neurosci.* **20**, RC89 (2000).
21. Conrad, K. L. et al. Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving. *Nature* **454**, 118–121 (2008).
22. Ambroggi, F. et al. Stress and addiction: glucocorticoid receptor in dopaminergic neurons facilitates cocaine seeking. *Nature Neurosci.* **12**, 247–249 (2009).
23. Liu, S. J. & Zukin, R. S. Ca²⁺-permeable AMPA receptors in synaptic plasticity and neuronal death. *Trends Neurosci.* **30**, 126–134 (2007).
24. Britt, J. P. et al. Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. *Neuron* **76**, 790–803 (2012).
25. French, S. J. & Totterdell, S. Hippocampal and prefrontal cortical inputs monosynaptically converge with individual projection neurons of the nucleus accumbens. *J. Comp. Neurol.* **446**, 151–165 (2002).
26. French, S. J. & Totterdell, S. Individual nucleus accumbens-projection neurons receive both basolateral amygdala and ventral subicular afferents in rats. *Neuroscience* **119**, 19–31 (2003).
27. Goto, Y. & Grace, A. A. Dopamine-dependent interactions between limbic and prefrontal cortical plasticity in the nucleus accumbens: disruption by cocaine sensitization. *Neuron* **47**, 255–266 (2005).
28. Pascoli, V., Turiault, M. & Lüscher, C. Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. *Nature* **481**, 71–75 (2011).
29. Robbe, D., Kopf, M., Remaury, A., Bockaert, J. & Manzoni, O. J. Endogenous cannabinoids mediate long-term synaptic depression in the nucleus accumbens. *Proc. Natl Acad. Sci. USA* **99**, 8384–8388 (2002).
30. Kasanetz, F. et al. Transition to addiction is associated with a persistent impairment in synaptic plasticity. *Science* **328**, 1709–1712 (2010).
31. Lüscher, C. & Huber, K. M. Group 1 mGluR-dependent synaptic long-term depression: mechanisms and implications for circuitry and disease. *Neuron* **65**, 445–459 (2010).
32. Clem, R. L. & Huganir, R. L. Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. *Science* **330**, 1108–1112 (2010).
33. McCutcheon, J. E. et al. Group I mGluR activation reverses cocaine-induced accumulation of calcium-permeable AMPA receptors in nucleus accumbens synapses via a protein kinase C-dependent mechanism. *J. Neurosci.* **31**, 14536–14541 (2011).
34. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
35. Lüscher, C. & Malenka, R. C. NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD). *Cold Spring Harb. Perspect. Biol.* **4**, a005710 (2012).
36. Counotte, D. S., Schiefer, C., Shaham, Y. & O'Donnell, P. Time-dependent decreases in nucleus accumbens AMPA/NMDA ratio and incubation of sucrose craving in adolescent and adult rats. *Psychopharmacology* **231**, 1675–1684 (2013).
37. Smith, D. G. & Robbins, T. W. The neurobiological underpinnings of obesity and binge eating: a rationale for adopting the food addiction model. *Biol. Psych.* **9**, 804–810 (2013).
38. Bock, R. et al. Strengthening the accumbal indirect pathway promotes resilience to compulsive cocaine use. *Nature Neurosci.* **16**, 632–638 (2013).
39. Lee, B. R. et al. Maturation of silent synapses in amygdala-accumbens projection contributes to incubation of cocaine craving. *Nature Neurosci.* **16**, 1644–1651 (2013).
40. Kombian, S. B. & Malenka, R. C. Simultaneous LTP of non-NMDA- and LTD of NMDA-receptor-mediated responses in the nucleus accumbens. *Nature* **368**, 242–246 (1994).
41. Grueter, B. A., Brasnjo, G. & Malenka, R. C. Postsynaptic TRPV1 triggers cell type-specific long-term depression in the nucleus accumbens. *Nature Neurosci.* **13**, 1519–1525 (2010).
42. Bellone, C. & Lüscher, C. Cocaine triggered AMPA receptor redistribution is reversed *in vivo* by mGluR-dependent long-term depression. *Nature Neurosci.* **9**, 636–641 (2006).
43. Koya, E. et al. Role of ventral medial prefrontal cortex in incubation of cocaine craving. *Neuropharmacology* **56** (suppl. 1), 177–185 (2009).

Acknowledgements We thank D. Huber and M. Brown as well the members of the Lüscher laboratory for discussion and comments on the manuscript, and A. Hiver for assisting in animal surgery. This work was financed by a grant from the Swiss National Science Foundation, the National Center of Competence in Research (NCCR) 'SYNAPSY - The Synaptic Bases of Mental Diseases' of the Swiss National Science Foundation and a European Research Council advanced grant (MeSSI). J.T. is supported by an MD-PhD grant of the Swiss Confederation. The E.V. laboratory was supported by an Avenir grant (Inserm) and by the Agence Nationale de la Recherche.

Author Contributions V.P. performed all the *in vitro* electrophysiological recordings. J.T. performed all surgery and behavioural experiments with the assistance of E.C.O'C. J.E. and E.V. performed the fluorescence immunohistochemical experiments. C.L., P.V., J.T. and E.C.O'C. designed the study, and C.L. wrote the manuscript with the help of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.L. (christian.luscher@unige.ch).

METHODS

Animals. Mice were C57BL/6 or heterozygous BAC transgenic mice in which tomato expression was driven by D1R (Drd1a-tdTomato from Jackson Laboratories) gene regulatory elements or eGFP driven by D2R (Drd2-eGFP from GENSAT) gene regulatory elements. Males and females were used. Transgenic mice had been backcrossed in the C57BL/6 line for a minimum of four generations. Mice were single housed after surgery. All animals were kept in a temperature- and humidity-controlled environment with a 12 h light/12 h dark cycle (lights on at 7:00). All procedures were approved by the Institutional Animal Care and Use Committee of the University of Geneva.

Stereotaxic injections. AAV1-CAG-ChR2-eYFP (eYFP; also called venus) or AAV1-CAG-eYFP (for control mice) produced at the University of North Carolina (Vector Core Facility) was injected into the mPFC (focusing on the infralimbic area), vHipp or BLA of 5- to 6-week-old mice. Anaesthesia was induced at 5% and maintained at 2.5% isoflurane (w/v) (Baxter AG) during the surgery. The animal was placed in a stereotaxic frame (Angle One) and craniotomies were performed using stereotaxic coordinates adapted from a mouse brain atlas⁴⁴ (for mPFC: anterior-posterior = +1.9; medial-lateral = \pm 0.3; dorsal-ventral (from the surface of the brain) = -2.5; for vHipp: anterior-posterior = -3.5; medial-lateral = \pm 2.8; dorsal-ventral = -4.2; for BLA: anterior-posterior = -1.4; medial-lateral = \pm 3.0; dorsal-ventral = -3.8). Injections of virus (0.5 μ l per injection site) used graduated pipettes (Drummond Scientific Company), broken back to a tip diameter of 10–15 μ m, at an infusion rate of \sim 0.05 μ l min⁻¹. Injections of ChR2 were made a minimum of 1 week before the self-administration surgery, and thus optogenetic manipulations typically occurred no sooner than 7 weeks after the injection. For anterograde tracing studies, 6- to 8-week-old C57BL/6J mice were infected with both AAV5-EF1 α -mCherry in the mPFC and AAV5-EF1 α -eYFP in the vHipp to allow observation of multiple inputs into NAc.

For retrograde tract-tracing studies, mice were stereotaxically injected with the retrograde tracer, cholera toxin subunit B conjugated to a fluorescent marker (cholera toxin subunit B, Alexa Fluor 594 conjugate, Molecular Probes). Microinjection needles were placed into the NAc medio-dorsal (anterior-posterior = +1.18; medial-lateral = +0.5; dorsal-ventral = -3.7) or NAc medio-ventral (anterior-posterior = +1.18; medial-lateral = +0.5; dorsal-ventral = -4.2) and 0.5 μ l was injected over 5 min. The injector was left in place for an extra 5 min to allow for diffusion of cholera toxin subunit B particles away from the injection site. Animals returned to their home cages for 14 days before tissue preparation.

Tissue preparation and immunofluorescence. Mice were rapidly anaesthetized with pentobarbital (500 mg kg⁻¹, intraperitoneally, Sanofi-Aventis) and transcardially perfused with 4% (w/v) paraformaldehyde in 0.1 M sodium phosphate buffer (pH 7.5) (ref. 45). Brains were post-fixed overnight in the same solution and stored at 4 °C. Thirty-micrometre-thick sections were cut with a vibratome (Leica) and stored at -20 °C in a solution containing 30% (v/v) ethylene glycol, 30% (v/v) glycerol, and 0.1 M sodium phosphate buffer, until they were processed for immunofluorescence. Sections were processed as follows. Free-floating sections were rinsed in Tris-buffered saline (TBS: 0.25 M Tris and 0.5 M NaCl, pH 7.5), incubated for 5 min in TBS containing 3% H₂O₂ and 10% methanol (v/v), and then rinsed three times 10 min in TBS. After incubation for 15 min in 0.2% (v/v) Triton X-100 in TBS, sections were rinsed three times in TBS again. Sections were then incubated for 1 h in a solution of BSA 3% in TBS. Finally, they were incubated 72 h at 4 °C with the primary antibodies mouse anti-DARPP-32 (1:1,000, a gift from P. Greengard, mouse monoclonal clone C54 (ref. 46)) and rabbit anti-RFP (1:1,000, Medical & Biological Laboratories, rabbit polyclonal, lot 042). After incubation with the primary antibodies, sections were rinsed three times for 10 min in TBS and incubated for 45 min with goat Cy3- and Cy5-coupled secondary antibodies (1:400, Jackson ImmunoResearch, lot 114787 and Life Technologies 1305923). Sections were rinsed for 10 min twice in TBS and twice in tris-buffer (0.25 M Tris) before mounting in 1,4-diazabicyclo-[2.2.2]-octane (DABCO, Sigma-Aldrich). Double- or triple-labelled images from each region of interest were obtained using sequential laser scanning confocal microscopy (Zeiss LSM510 META). Photomicrographs were obtained with the following band-pass and long-pass filter settings: GFP (band-pass filter: 505–530), Cy3 (band-pass filter: 560–615) and Cy5 (long-pass filter 650). The objectives and the pinhole setting (1 airy unit) remained unchanged during the acquisition of a series for all images. The thickness of the optical section was \sim 1.6 μ m with a \times 20 objective and \sim 6 μ m with a \times 10 objective. GFP-labelled neurons were pseudo-coloured green and other immunoreactive neurons were pseudo-coloured red or blue. For images showing infected structure with AAV1-CAG-ChR2-eYFP, 30 μ m-thick coronal sections were cut with a vibratome (Leica), stained with Hoechst (Sigma-Aldrich) and mounted with Mowiol (Sigma-Aldrich). Full images of brain slices from mPFC, BLA and vHipp were obtained with a MIRAX (Carl Zeiss) system equipped with a Plan-Apochromat \times 20/0.8 objective, together with 4',6-diamidino-2-phenylindole (DAPI) (emission: 455/50) and fluorescein isothiocyanate (FITC) (emission: 515–565) filters.

Implantation of jugular vein catheter. The surgical procedure was adapted from refs 47, 48. Mice were anaesthetized with a mix of ketamine (60 mg kg⁻¹, Ketalar) and xylazine (12 mg kg⁻¹, Rompun) solution. Catheters (CamCaths, model MIVSA) made of silicone elastomer tubing (outside diameter 0.63 mm, inside diameter 0.30 mm) were inserted 1.0–1.2 cm in the right jugular vein, about 5 mm from the pectoral muscle, to reach the right atrium. The other extremity of the catheter was placed subcutaneously in the mid-scapular region and connected to stainless steel tubing appearing outside the skin. Blood reflux in the tubing was checked to confirm correct placement of the catheter. Mice were allowed to recover for 3–5 days before the start of drug self-administration training and received antibiotics (Baytril 10%, 1 ml in 250 ml of water) in the drinking water during this period. Catheters were flushed daily with a heparin solution (Heparin 'Bichsel' in saline (30 IU) during the recovery period and just before and after each self-administration session.

Self-administration apparatus. All behavioural experiments were performed during the light phase and took place in mouse operant chambers (ENV-307A-CT, Med Associates) situated in sound-attenuating cubicle (Med Associates). Two retractable levers were present on both sides of one wall of the chamber and a food pellet dispenser was also present only for procedures involving food delivery. A cue-light was located above each lever and a house light was present in each chamber. During intravenous drug self-administration sessions, the stainless steel tubing of the catheter device was connected through a CoEx PE/PVC tubing (BCOEX-T25, Instech Solomon) to a swivel (Instech Solomon) and then an infusion pump (PHM-100, Med-Associates). The apparatus was controlled and data captured using a PC running MED-PC IV (Med-Associated).

Drug self-administration acquisition. Mice were deprived of food for 12 h before the first self-administration session to promote exploratory activity. No food shaping was undertaken and mice were given food access *ad libitum* after the first session. Each session was 120 min in duration and started with the illumination of the house light and the insertion of the two levers into the operant chamber. During the first five sessions, a single press on the active lever (termed fixed-ratio one, or FR1) resulted in an infusion of 0.75 mg kg⁻¹ of cocaine (cocaine hydrochloride, provided by the pharmacy of Geneva University Hospital, dissolved in 0.9% saline at 0.75 mg ml⁻¹ and delivered at 0.0177 ml s⁻¹ as a unit dose depending on the weight of the mouse) paired with a 5 s continuous illumination of the cue light above the active lever. Completion of the fixed-ratio also initiated a timeout period of 20 s during which cocaine was no longer available. For the next five sessions, two lever presses were needed to activate the infusion pump (FR2). The active lever (left or right lever) was randomly assigned for each mouse. To avoid an overdose of cocaine, a maximum of 45 infusions were allowed per session. Only mice having reached a stable rate of at least 70% of correct lever responses (number of active lever responses divided by total lever responses over the three last sessions of acquisition) were included in the study. Saline control mice undertook the same procedure as cocaine mice except that saline (NaCl 0.9% B. Braun) replaced cocaine infusions. After acquisition, mice were randomly attributed to either behaviour or electrophysiology experiments (except wild-type mice, which were only used in behaviour experiments). Recordings were never performed after cue-associated seeking tests to avoid confounding effects on synaptic transmission⁴⁹.

Optic fibre cannulation and *in vivo* optogenetic stimulation protocols. After completing the acquisition of self-administration, all mice were put in forced abstinence for 30 days in their home cage. At days 15–20, mice destined for seeking tests or *ex vivo* validation were bilaterally implanted with a chronically indwelling optic-fibre cannula (made in house as described in ref. 50) into the NAc (anterior-posterior = +1.5; medial-lateral = \pm 1.6; dorsal-ventral = -3.9 with a 10° angle) using stereotaxic apparatus as described above. Two screws were fixed into the skull to support the implant, which was further secured with dental cement.

DPSS blue light lasers (MBL-473/50 mW; CNI Lasers) connected to the indwelling fibre optic by customized patch cords (BFL37-200 Custom; Thor Labs) and a double rotary joint (FRJ-1X2i; Doric Lenses) allowed mice to move freely during stimulation. The laser was triggered to deliver 4 ms pulses at 1 Hz or 13 Hz for 10 min, and the protocol was applied in the home cage 4 h before the first drug-seeking session or killing for *ex vivo* electrophysiology recordings. Mice from the control group were, in a randomized manner, either not infected with ChR2 or infected with eYFP only and received one of the optogenetic protocols, or were infected with ChR2 but were not exposed to light-stimulation. Seeking behaviour did not differ among these three control conditions, so the data were collapsed.

Test of cue-associated drug seeking. Thirty days after the final self-administration session (that is, day 40), mice were assigned to optogenetic protocol groups (control, 1 or 13 Hz) according to performance during self-administration acquisition, such that acquisition did not differ between the groups. Control animals were included for each batch of tested animals explaining the greater number of mice in the control group. The cue-associated drug-seeking test was a 60 min session, identical to the cocaine acquisition period (house light on, insertion of the two levers), except that one press on the active lever (FR1 schedule) now triggered illumination of the cue

light for 5 s but without a cocaine infusion or a timeout period. The infusion pump was also activated during the drug-seeking session, because the pump noise provided an extra drug-associated cue. One week later (that is, day 48), mice undertook a second drug-seeking session (with no further optogenetic stimulation 4 h before) to assess the persistence of the optogenetic protocol on seeking behaviour. At the end of any behavioural experiment, mice were euthanized and brains fixed in paraformaldehyde to prepare histological slices for verification of Chr2 expression and cannula placement (see Extended Data Fig. 9).

Food training after drug seeking. Immediately after the second drug-seeking test, subgroups of mice were food deprived for 12 h and then started nine consecutive sessions of operant training for food to examine whether optogenetic protocols had long-term effects on the acquisition of new learning for a natural reward. The test was performed in similar chambers as used previously (ENV-307A-CT, Med Associates), except that two nose-poke holes were present instead of levers at each side of a food pellet dispenser. One nose-poke in the active hole (randomly assigned) illuminated the cue light for 5 s and triggered the delivery of one food pellet (50% sucrose, AIN-76A Rodent Tablet 20 mg, Testdiet). The response requirement was set at FR1 for sessions 1–3, FR2 for sessions 4–6 and FR3 for sessions 7–9. The house light was turned off during the entire session. Each session ended after 30 min or when 45 pellets had been delivered, whichever occurred first. Each mouse received 2.7–3 g per day of standard laboratory chow during the food training procedure to maintain a stable body weight without any further weight gain.

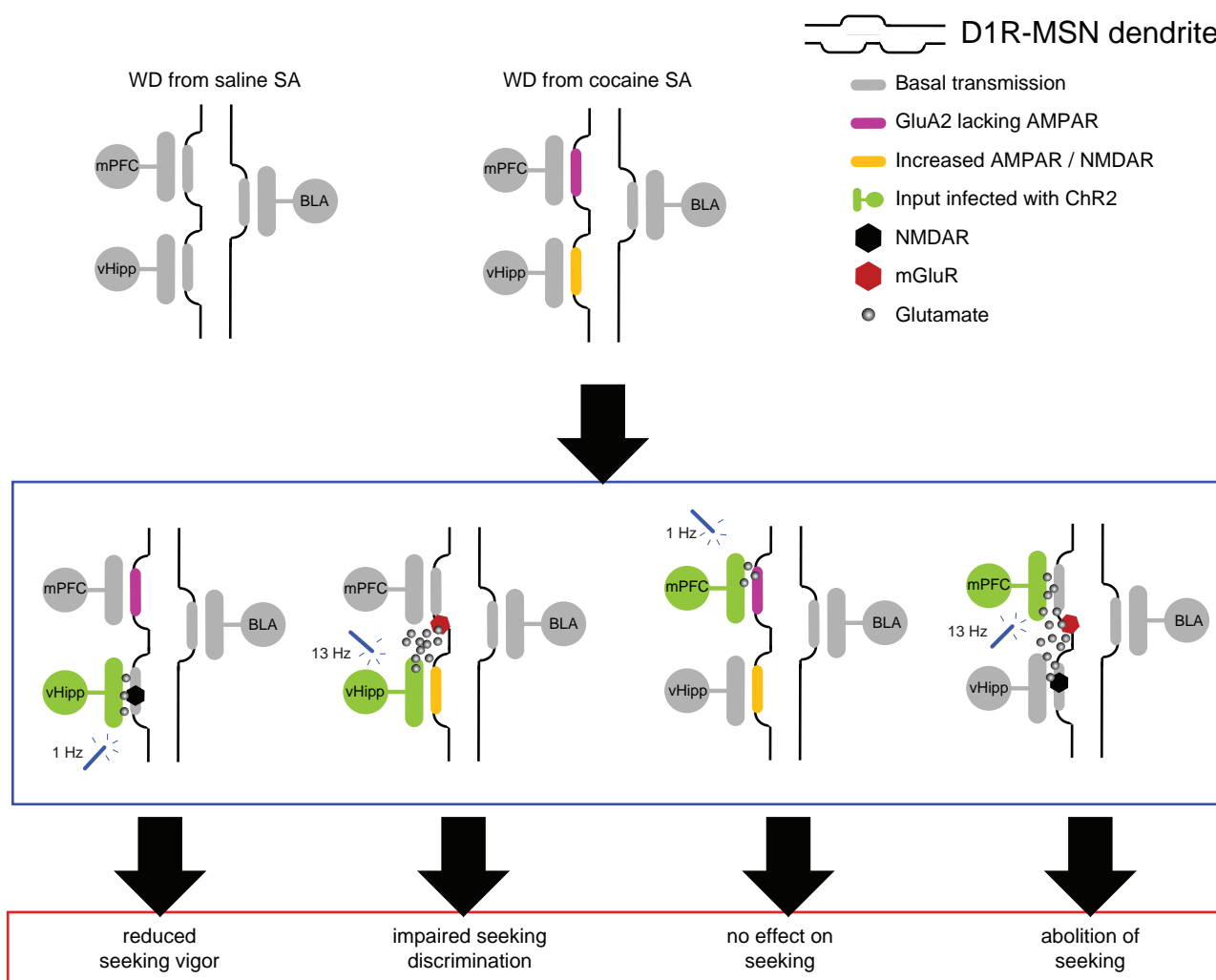
Effect of optogenetic protocols on cue-associated food seeking. To assess the specificity of the optogenetic treatment for drug versus food cue-associated seeking, we also tested a new cohort of mice for cue-associated food seeking after the optogenetic protocol. Mice underwent food training in the same conditions as for the cocaine self-administration group except that cocaine infusions were replaced by sucrose pellet delivery. Sessions ended after 30 min or when 45 pellets were delivered, whichever occurred first. After implantation of fibre optic cannula at day 15–20 of forced abstinence from sucrose pellet consumption, mice were assigned to the control or 13 Hz groups (such that acquisition did not differ between the groups) and received the optogenetic treatment 4 h before a test for food-reward seeking that was undertaken in the same way as the test for cocaine seeking. The sound of the pellet dispenser was also paired with the food cue light during the session. Food seeking sessions lasted 30 min. Mice received 2.7–3 g per day of standard laboratory chow over the entire procedure to maintain a stable body weight (no weight gain).

Slice electrophysiology. Coronal 200–250 μ m slices of mouse brain were prepared in cooled artificial cerebrospinal fluid containing (in mM): NaCl 119, KCl 2.5, MgCl 1.3, CaCl₂ 2.5, Na₂HPO₄ 1.0, NaHCO₃ 26.2 and glucose 11, bubbled with 95% O₂ and 5% CO₂. Slices were kept at 32–34 °C in a recording chamber superfused with 2.5 ml min⁻¹ artificial cerebrospinal fluid. Visualized whole-cell voltage-clamp recording techniques were used to measure holding and synaptic responses of MSNs of the NAc shell, identified by the presence of the eGFP or td-Tomato of BAC transgenic mice by using a fluorescent microscope (Olympus BX50WI, fluorescent light U-RFL-T). Note that when using Chr2-eYFP in BAC transgenic mice, D2R-MSNs were identified as D2R-eGFP-positive neurons in Drd2-eGFP mice, which could be visualized against fluorescence from Chr2-eYFP, or as tomato-negative neurons in Drd1a-tdTomato mice. D1R-MSNs were always identified as tomato-positive cells. This approach was validated in a previous report (ref. 28). The holding potential was -70 mV, and the access resistance was monitored by a hyperpolarizing step of -14 mV with each sweep, every 10 s. The liquid junction potential was small (-3 mV), and therefore traces were not corrected. Experiments were discarded if the access resistance varied by more than 20%. Currents were amplified (Multiclamp 700B, Axon Instruments), filtered at 5 kHz and digitized at 20 kHz (National Instruments Board PCI-MIO-16E4, Igor, WaveMetrics). All experiments were performed in the presence of picrotoxin (100 μ M).

For recordings of light- and electrically evoked EPSCs, the internal solution contained (in mM) 130 CsCl, 4 NaCl, 5 creatine phosphate, 2 MgCl₂, 2 Na₂ATP, 0.6 Na₃GTP, 1.1 EGTA, 5 HEPES and 0.1 mM spermine. In some cases, QX-314 (5 mM) was added to the solution to prevent action currents. Synaptic currents were electrically evoked by stimuli (50–100 μ s) at 0.1 Hz through bipolar stainless steel electrode placed onto the tissue. For optogenetic experiments, light-EPSCs were evoked with 4 ms blue light pulses from an optic fibre directed at the NAc shell that was coupled to a DPSS blue light laser. Low-frequency stimulation (1 or 13 Hz for 10 min) was applied with 4 ms light pulses and the magnitude of LTD was determined by comparing average EPSCs that were recorded 20–30 min after induction to EPSCs recorded immediately before induction. To isolate AMPAR-evoked EPSCs, the NMDA antagonist D-AP5 (50 μ M) was bath applied. The NMDAR component was calculated as the difference between the EPSCs measured in the absence and in the presence of D-AP5. The AMPAR/NMDAR ratio was calculated by dividing the peak amplitudes. The rectification index of AMPAR was calculated as the ratio of the chord conductance calculated at negative potential divided by chord conductance at positive potential. Note that, in example traces, stimulation artefacts were removed.

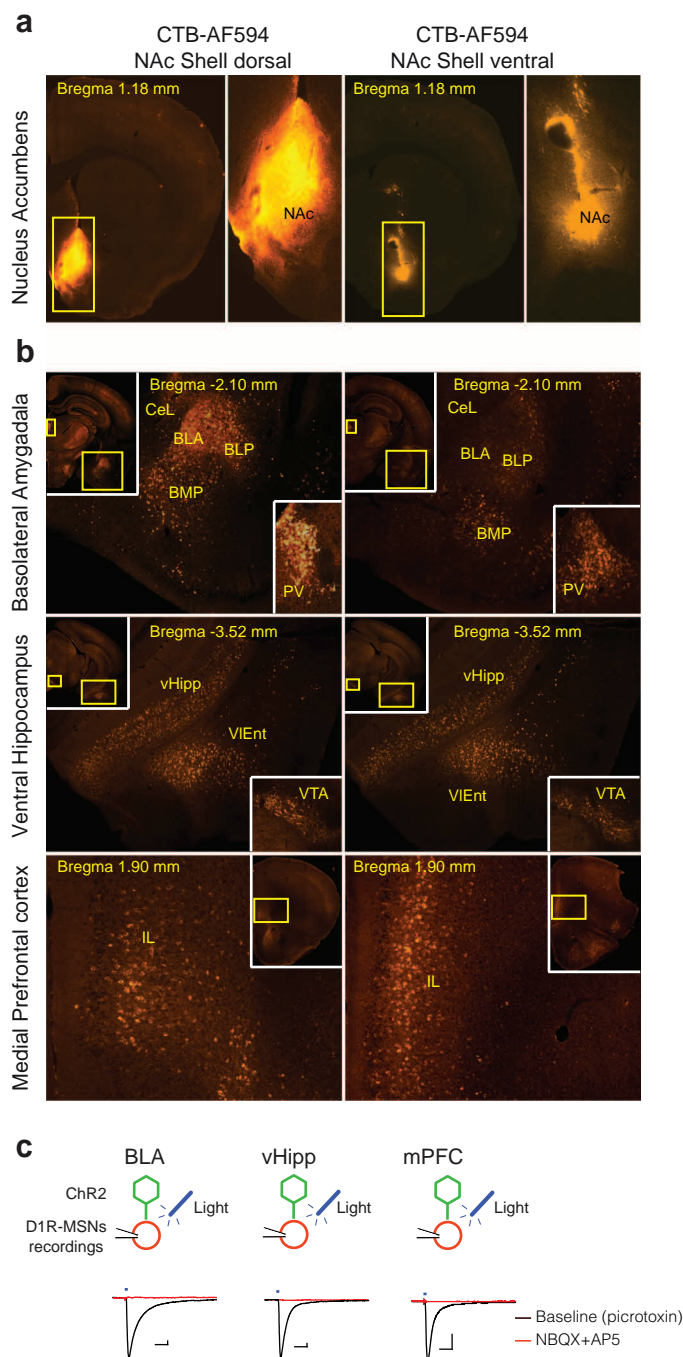
For recordings of mEPSCs the internal solution contained (in mM): 140 K Glu, 2 MgCl₂, 5 KCl, 0.2 EGTA, 10 HEPES, 4 Na₂ATP, 0.3 Na₃GTP and 10 creatine phosphate. Miniature EPSCs were recorded in the presence of tetrodotoxin (0.5 μ M). The frequencies, amplitudes and kinetic properties of these currents were then analysed using the Mini Analysis software package (version 4.3, Synaptosoft). Note that recordings were performed with the electrophysiologist blinded to the self-administration condition (cocaine or saline), although for practical reasons the experimenter was aware of other conditions (for example, Chr2 infected or not, etc.). **Statistics.** No statistics were used to determine group sample size; however, sample sizes were similar to those used in previous publications from our group and others reporting self-administration in mice. Multiple comparisons were first subject to mixed-factor ANOVA defining both between- (for example, D1R- or D2R-MSN cells; saline or cocaine self-administration groups; control, 1 Hz or 13 Hz protocols; BLA, vHIPP, mPFC inputs, etc.) and/or within- (for example, active or inactive levers) group factors. Where significant main effects or interaction terms were found ($P \leq 0.05$; or $P \leq 0.1$ indicative of a trend), further comparisons were made by a two-tailed Student's *t*-test with Bonferroni corrections applied when appropriate (that is, the level of significance equalled 0.05 divided by the number of comparisons). Single comparisons of between- or within-group measures were made by two-tailed non-paired or paired Student's *t*-test, respectively. ANOVAs for main figures are provided (see Extended Data Table 1a, b).

44. Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd edn (Elsevier Academic Press, 2008).
45. Bertran-Gonzalez, J. *et al.* Opposing patterns of signaling activation in dopamine D1 and D2 receptor-expressing striatal neurons in response to cocaine and haloperidol. *J. Neurosci.* **28**, 5671–5685 (2008).
46. Ouimet, C. C. *et al.* DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. *J. Neurosci.* **4**, 111–124 (1984).
47. Thomsen, M. & Caine, S. B. Intravenous drug self-administration in mice: practical considerations. *Behav. Genet.* **37**, 101–118 (2006).
48. Chistyakov, V. S. & Tsibulsky, V. L. How to achieve chronic intravenous drug self-administration in mice. *J. Pharmacol. Toxicol. Methods* **53**, 117–127 (2006).
49. Gipson, C. D. *et al.* Relapse induced by cues predicting cocaine depends on rapid, transient synaptic potentiation. *Neuron* **77**, 867–872 (2013).
50. Sparta, D. R. *et al.* Construction of implantable optical fibers for long-term optogenetic manipulation of neural circuits. *Nature Protocols* **7**, 12–23 (2011).



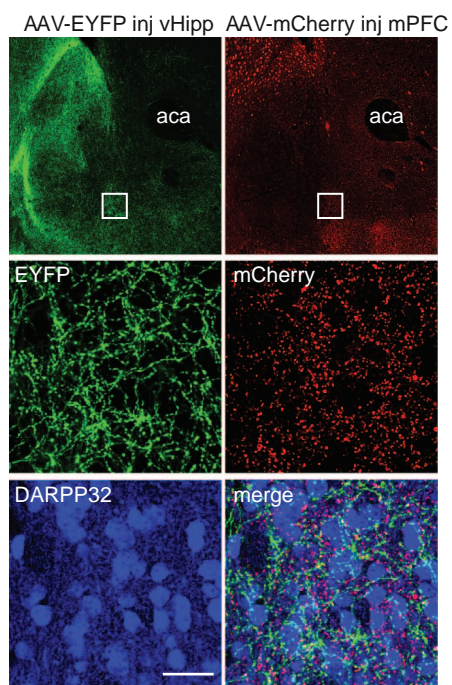
Extended Data Figure 1 | Graphical abstract. Top left: main excitatory afferents onto NAc shell D1R-MSNs (BLA: basolateral amygdala, vHipp: ventral subiculum of the hippocampus and mPFC: medial prefrontal cortex), which at baseline contain synapses that express NMDARs and GluA2-containing AMPARs. Top right: 1 month after withdrawal (WD) from cocaine self-administration (SA), mPFC synapses onto NAc shell D1R-MSNs express

GluA2-lacking AMPARs whereas more GluA2-containing AMPARs are added at vHipp synapses. Effects of NMDAR- or mGluR1-dependent (1 or 13 Hz, respectively) light protocols applied at specific inputs (shown in green) on cocaine-evoked plasticity are illustrated, together with the consequence for cue-associated seeking behaviour.

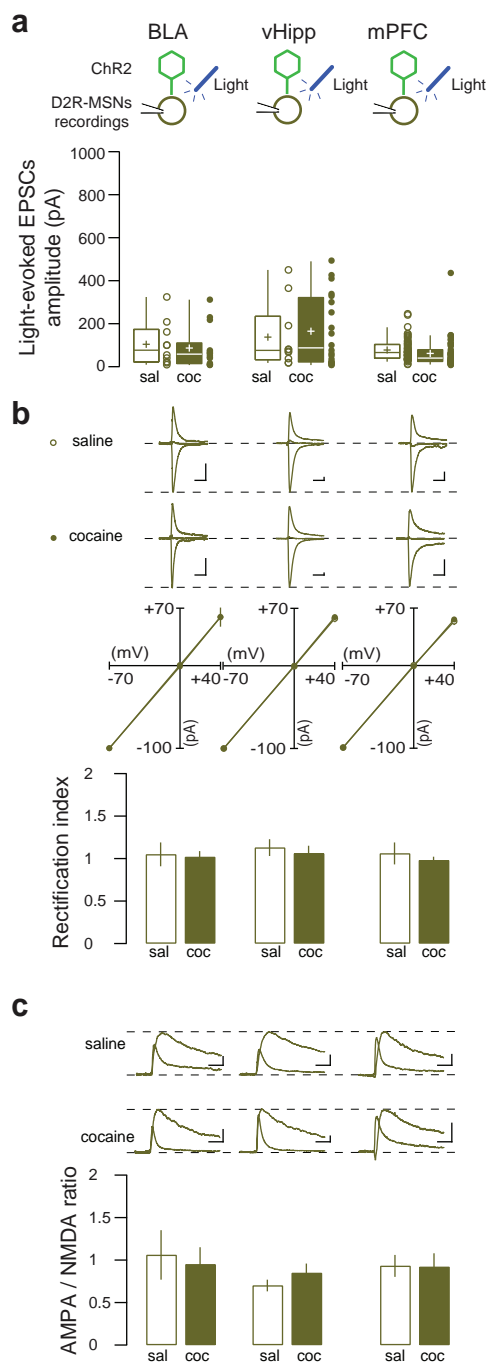


Extended Data Figure 2 | Identification and optogenetic targeting of excitatory inputs to the NAc shell. **a**, Retrograde labelling with cholera toxin subunit B (AF594) injected into the NAc shell. Confocal images of injection sites (top) in the medio-dorsal (left) and medio-ventral (right) NAc shell are shown, regions where electrophysiology recordings were performed.

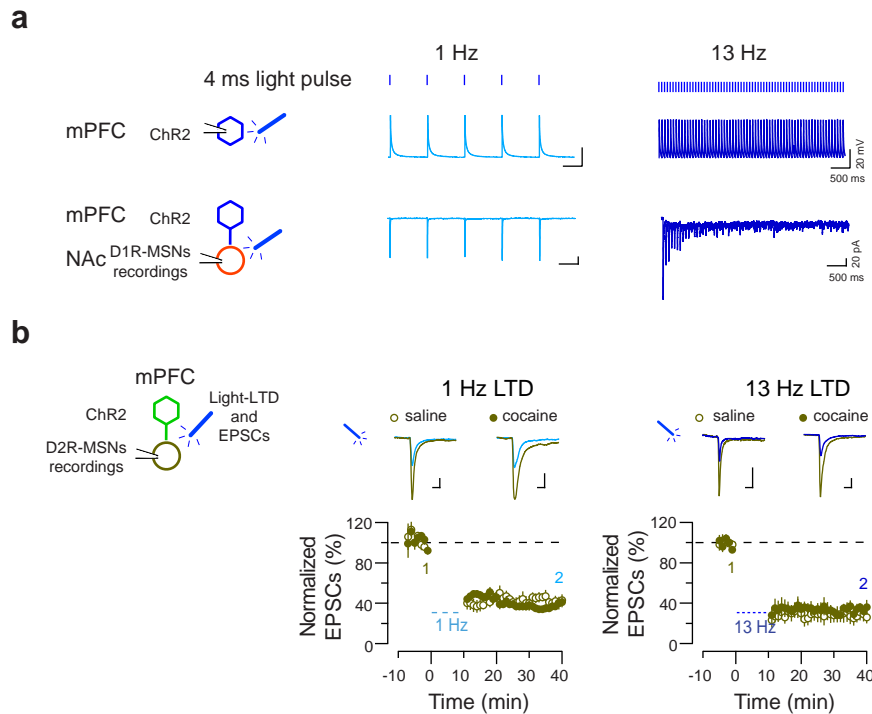
b, Labelled cell bodies in corresponding projection areas (basolateral amygdala, ventral subiculum of the hippocampus and medial prefrontal cortex) are shown, with no discernable segregation between the medio-dorsal or medio-ventral NAc shell. For each projection area, the insert shows a complete hemisphere coronal section together with a zoomed image of the region of interest (indicated by yellow box). IL, infralimbic; CeL; central amygdala lateral; BLP, basolateral amygdala posterior; BMP, basomedial amygdala posterior; PV, paraventricular thalamic nucleus; vHipp, central subiculum of the hippocampus; VIEnt, ventral intermediate entorhinal cortex; VTA, ventral tegmental area. **c**, Schematic of experiment (top) with light-evoked EPSCs recorded in D1R-MSNs of the NAc shell of mice infected with AAV1-ChR2-eYFP in the BLA (bottom left), vHipp (middle) or vmPFC (right) before and after bath application of glutamate receptor antagonists (NBQX 10 μ M and AP5 50 μ M for AMPAR and NMDAR, respectively). Scale bars, 20 ms, 50 pA.



Extended Data Figure 3 | Individual MSNs receive inputs from multiple projection areas. **a**, Confocal images of NAc from a mouse infected with AAV5-EF1-eYFP and AAV5-EF1-mCherry in the vHipp (left) and mPFC (right), respectively, at low magnification (first row). At higher magnification (second and third rows) eYFP from vHipp and mCherry from mPFC are present around MSNs stained by DARPP-32 (blue). aca, anterior commissure. Scale bar, 50 μ m.

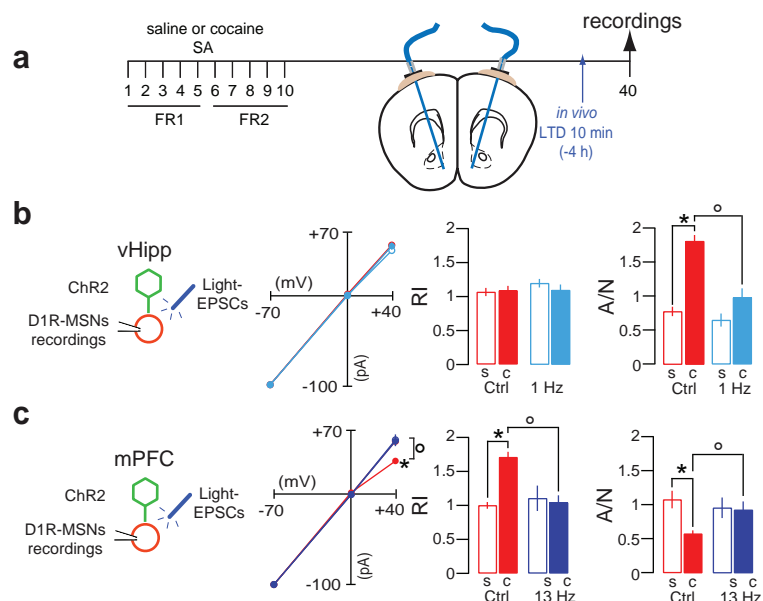


Extended Data Figure 4 | Cocaine self-administration does not evoke input-specific plasticity in D2R-MSNs. **a**, Top, schematic of whole-cell recordings of NAc shell D2R-MSNs of mice that 1 month previously self-administered saline (open points) or cocaine (filled points) and were infected with AAV1-ChR2-eYFP in the BLA (left), vHipp (middle) or mPFC (right). Bottom, after cocaine self-administration the mean amplitude of light-evoked EPSCs was not changed at any input onto D2R-MSNs (effect of group (saline/cocaine) and group \times input (BLA/vHipp/mPFC) all not significant) ($n = 10/14$ for BLA (saline/cocaine), $n = 10/20$ for vHipp and $n = 60/51$ for mPFC). **b**, For each input, the rectification index (RI) was calculated. Example traces are shown (top), with the I/V plot (middle) and group mean rectification index data (bottom). Cocaine did not modify normalized AMPAR-EPSCs at +40 mV from BLA, vHipp or mPFC inputs ($t_{12} = -0.20$, $P = 0.84$, $t_{18} = 0.44$, $P = 0.67$ and $t_{20} = 0.43$, $P = 0.67$, respectively). The rectification index was also unchanged at D2R-MSN synapses from BLA, vHipp or mPFC inputs ($t_{12} = -0.32$, $P = 0.75$, $t_{18} = -0.51$, $P = 0.62$ and $t_{20} = -0.67$, $P = 0.51$ respectively). Scale bars, 20 ms, 20 pA. **c**, For the same cells as shown in **b**, the A/N ratio was calculated. For each input, example traces are shown (top), with group mean A/N ratios (bottom). Cocaine did not alter the A/N ratio at inputs onto D2R-MSNs from the BLA, vHipp or mPFC ($t_{12} = -0.19$, $P = 0.85$, $t_{18} = 1.20$, $P = 0.25$ and $t_{20} = -0.04$, $P = 0.97$, respectively). Scale bars, 20 ms, 20 pA. Error bars, s.e.m.



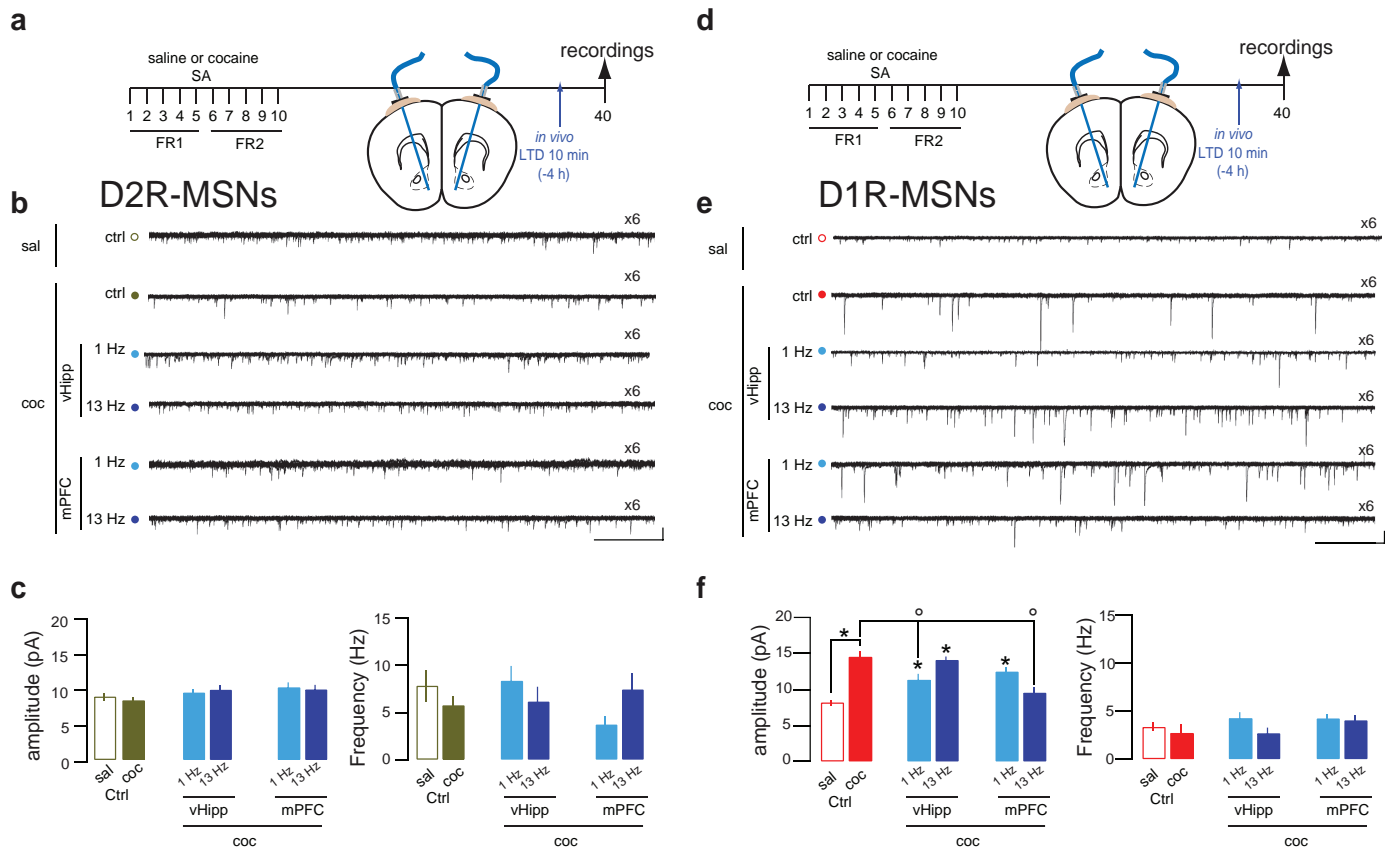
Extended Data Figure 5 | mPFC and NAc recordings during 1 and 13 Hz optogenetic protocols and LTD in NAc D2R-MSNs induced by mPFC protocols applied after saline and cocaine self-administration. **a**, Right, schematic of whole-cell recordings in the mPFC or NAc from mice infected with ChR2 in mPFC. Top, light-evoked action potentials recorded in current clamp of ChR2-infected mPFC neurons and EPSCs recorded in voltage-clamp of D1R-MSNs (bottom) during the beginning of the 1 Hz (left) or 13 Hz (right) stimulation protocols. Note that EPSCs fail to follow the 13 Hz protocol. **b**, Top left, schematic of experiment. Bottom, graph of normalized light-evoked

EPSCs across time recorded in NAc D2R-MSNs from saline and mice that self-administered cocaine (each point represents mean of six sweeps), together with example traces (mean of 20 sweeps) before (1) and after (2) a 1 (left) or 13 Hz (right) light protocol was applied *ex vivo* (4 ms pulses at 1 or 13 Hz, 10 min). One month after saline or cocaine self-administration, the 1 Hz and 13 Hz protocols induced comparable LTDs in both groups (for 1 Hz: $50 \pm 4.9\%$ to $40 \pm 2.6\%$, Student's t -test $t_{14} = -1.86$, $P = 0.080$; $n = 7-9$ cells; for 13 Hz: $28 \pm 3.7\%$ to $33 \pm 4.2\%$, Student's t -test $t_{18} = 1.41$, $P = 0.18$; $n = 11-9$ cells). Error bars, s.e.m.



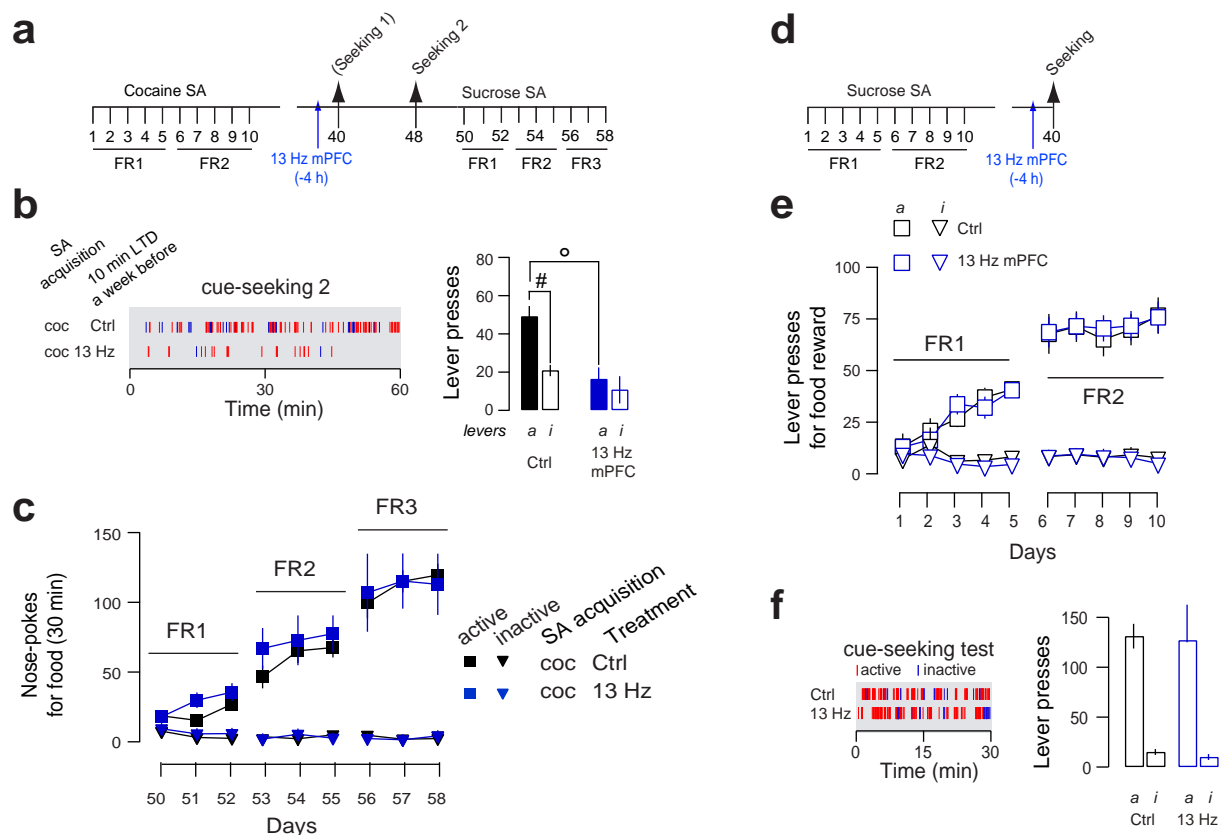
Extended Data Figure 6 | Optogenetic protocols applied *in vivo* reverse cocaine evoked-plasticity at NAc D1R-MSNs. **a**, Top, schematic of experiment. Mice were infected with ChR2 in the vHipp or mPFC and trained in saline or cocaine self-administration. Mice were then implanted with fibre optics targeting the NAc shell, and optogenetic protocols were applied *in vivo* 1 month after withdrawal. Four hours later, acute brain slices were prepared to assess the efficiency of optogenetic protocols applied *in vivo* to reverse cocaine-evoked plasticity at NAc D1R-MSNs. In brief, consistent with the *ex vivo* validation, the 1 Hz vHipp protocol applied *in vivo* normalized the *A/N* ratio, whereas the 13 Hz mPFC protocol normalized the *I/V* curve and rectification index in mice that self-administered cocaine. **b**, Left, schematic of experiment indicating that mice were infected with ChR2 in the vHipp. The rectification index and *A/N* ratio were determined with light-evoked EPSCs as described previously (see Fig. 2). The *I/V* plot is shown, together with group mean rectification index and *A/N* data. Cocaine did not alter normalized AMPAR-EPSCs recorded at +40 mV or the rectification index from vHipp inputs (same data as Fig. 3). The 1 Hz protocol applied *in vivo* was without effect on either of these measures in mice that self-administered saline or

cocaine (effect of group (saline versus cocaine), protocol (control versus 1 Hz) and group \times protocol, all not significant). The *A/N* ratio was increased at vHipp inputs in mice that self-administered cocaine (same data as Fig. 3), an effect that was reduced after the *in vivo* 1 Hz protocol (planned comparison, after ANOVA, by *t*-test, $t_{26} = 4.97$, $^{\circ}P < 0.001$) ($n = 13/15$ for 1 Hz saline/cocaine group). **c**, As for **b**, except that mice were infected with ChR2 in the mPFC. Cocaine decreased AMPAR-EPSCs recorded at +40 mV and increased the rectification index from mPFC inputs (same data as Fig. 3). The 13 Hz protocol applied *in vivo* increased AMPAR-EPSCs at +40 mV in mice that self-administered cocaine (planned comparison with cocaine control group, after ANOVA, by *t*-test, $t_{20} = 3.6$, $^{\circ}P < 0.01$) and decreased the rectification index in mice that self-administered cocaine (planned comparison with cocaine controls, after ANOVA, by *t*-test, $t_{20} = 5.2$, $^{\circ}P < 0.001$). The *A/N* ratio was decreased at mPFC inputs in mice that self-administered cocaine (same data as Fig. 3), an effect that was normalized after the *in vivo* 13 Hz protocol (planned comparison with cocaine controls, after ANOVA, by *t*-test, $t_{20} = 2.8$, $^{\circ}P = 0.01$) ($n = 11/9$ for 13 Hz saline/cocaine group). Error bars, s.e.m.



Extended Data Figure 7 | Assessing effects of *in vivo* light stimulation on mEPSCs recorded in D2R- and D1R-MSNs. **a**, Top, schematic of experiment. Mice were infected with Chr2 in the vHipp or mPFC and trained in saline or cocaine self-administration. Mice were then implanted with fibre optics targeting the NAc shell and optogenetic protocols applied *in vivo* 1 month after withdrawal. Four hours later, acute brain slices were prepared to assess the effect of optogenetic protocols applied *in vivo* on global excitatory transmission by recording mEPSCs at NAc D2R-MSNs. In brief, recordings from D2R-MSNs showed that mEPSCs were not affected by cocaine and not depressed by optogenetic LTD protocols applied *in vivo*. Note that although optogenetic protocols efficiently induced LTD at single inputs onto D2R-MSNs (Extended Data Fig. 5), this was not reflected by a decrease in mEPSC amplitudes. This may be accounted for by a presynaptic expression mechanism of LTD or that baseline amplitudes were already low such that a further depression only at a single input could not be measured by mEPSCs (that is, floor effect), which reflects synaptic transmission from multiple inputs. **b**, Example of mEPSCs recorded *ex vivo* in NAc shell D2R-MSNs in the presence of picrotoxin (100 μ M) and tetrodotoxin (0.5 μ M) (sample traces comprising six superimposed, 4 s traces). Scale bars, 20 pA, 500 ms. **c**, Histograms of group mean data of D2R-MSN mEPSC amplitudes (left) and frequency (right) are shown in control saline (sal) or cocaine (coc) conditions and after application of 1 Hz or 13 Hz light protocols at vHipp or mPFC synapses. Mean mEPSC amplitudes and frequencies were not changed by cocaine, and were not significantly decreased by protocols applied at either

vHipp or mPFC inputs ($n = 5$ to 13 cells per group). Error bars, s.e.m. **d**, **e**, As for **a** and **b** except in D1R-MSNs. In brief, the frequency of mEPSCs was not affected by cocaine self-administration or laser protocols. In contrast, in mice that self-administered cocaine the amplitude of mEPSCs was significantly larger than controls, in line with a postsynaptic expression mechanism. Protocols that were most efficient at restoring the A/N ratio at vHipp synapses when assessed on slice, namely the 13 Hz mPFC and the 1 Hz vHipp protocol, were also most efficient at restoring baseline mEPSC amplitudes when applied *in vivo*. This suggests that cocaine-evoked plasticity at vHipp inputs largely accounts for the observed increase in mEPSC amplitudes. **f**, Mean mEPSC amplitudes were increased by cocaine self-administration (versus saline control, by *t*-test, $t_{18} = 7.13$, $^*P < 0.001$). Protocols applied at vHipp terminals altered mEPSC amplitudes (one-way ANOVA comparing cocaine control, 1 and 13 Hz vHipp protocols: effect of protocol $F_{2,28} = 5.7$, $P < 0.01$). The 1 Hz but not the 13 Hz vHipp protocol reduced mEPSC amplitudes (versus cocaine control, for 1 Hz: $t_{18} = 2.8$, $^*P = 0.01$), although amplitudes remained significantly higher than saline control mice after either protocol (all $P < 0.01$). Protocols applied at mPFC terminals also altered mEPSC amplitudes (effect of protocol, $F_{2,29} = 13.1$, $P < 0.001$), and the 13 Hz but not 1 Hz mPFC protocol reduced mEPSC amplitudes (versus cocaine control, for 1 Hz: $t_{19} = 4.7$, $^*P < 0.001$). Moreover, amplitudes after the 13 Hz mPFC protocol did not differ from saline controls ($P = 0.14$). The frequencies of mEPSCs were not altered by cocaine or by protocols applied at either vHipp or mPFC inputs ($n = 8$ –11 cells per group). All error bars, s.e.m.

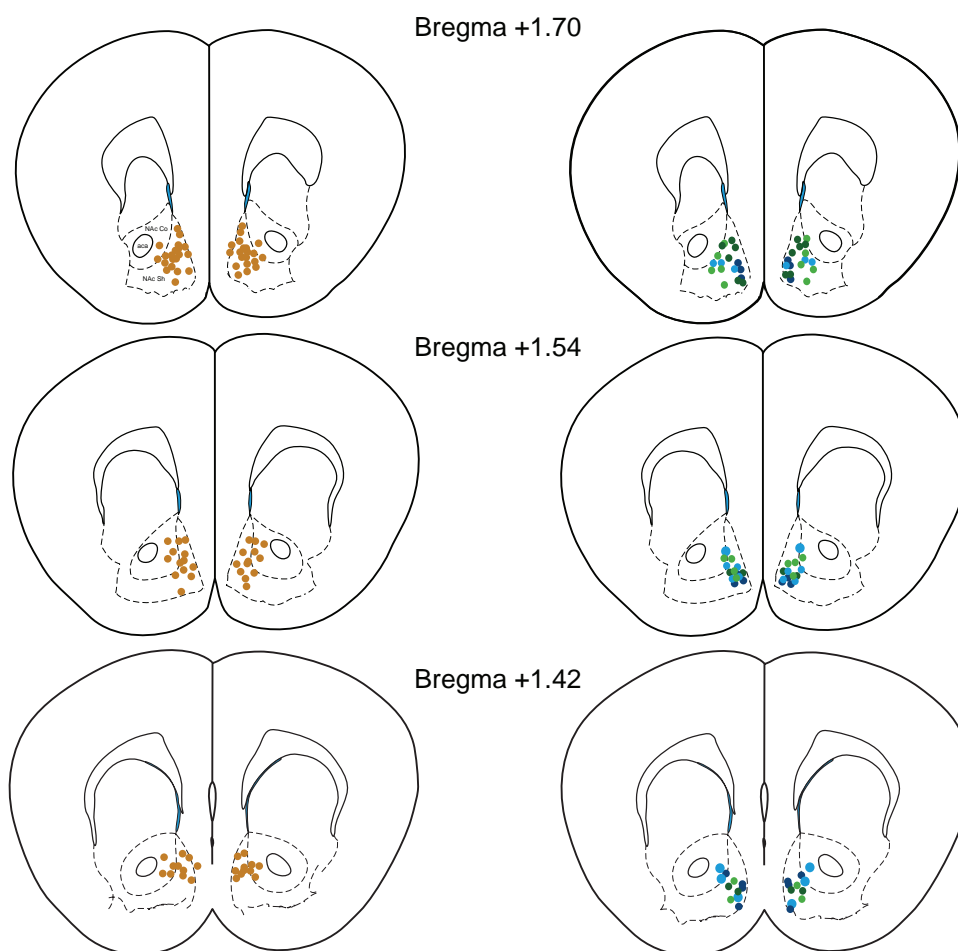


Extended Data Figure 8 | Optogenetic effect on cocaine seeking is persistent, and does not alter learning of a new response for food reward or food cue-associated seeking. **a**, Schematic of experiment. Cocaine control and 13 Hz mPFC group mice were exposed to a second cocaine cue-associated seeking test 1 week after the first test (see Fig. 5), but without further optogenetic intervention. Two days after this second test, subgroups of mice from each condition were trained to nose-poke for sucrose pellets during once-daily 30 min sessions over 9 days. **b**, Raster plots of cue-associated seeking behaviour (active and inactive lever presses) during the second test are shown (left), together with group mean data showing active (a: filled boxes) and inactive (i: open boxes) lever presses. Seeking was present in cocaine control mice during the second test, but was absent in the 13 Hz mPFC group (effect of group, group and lever group, all $P < 0.05$; active versus inactive lever comparison by t -test, $t_{43} = 6.13$, $^{\#}P < 0.001$ for controls $n = 44$ and $t_9 = 0.97$, $P = 0.35$ for 13 Hz mPFC, $n = 10$; active versus active lever comparison, $^{\circ}P < 0.01$). **c**, For the acquisition of a new response test, group mean values are shown of active and inactive nose-pokes across training days. Responding was

maintained by FR1, FR2 and FR3 schedules. ANOVA comparisons of nose-poke responding across 3 days at each fixed ratio schedule confirmed no difference between cocaine self-administration control ($n = 15$) and 13 Hz mPFC ($n = 4$) mice (effects of group and day \times group \times nose poke interactions, all not significant). **d**, Schematic of experiment. One month after food self-administration training, mice received the 13 Hz mPFC light protocol, 4 h before a 30 min test of seeking. **e**, Graph shows mean of active (a) and inactive (i) lever presses for sucrose pellets by new cohort of mice ($n = 14$) infected or not with ChR2 in the mPFC. Training parameters were identical to those used for the acquisition of cocaine self-administration, except that sessions lasted only 30 min. **f**, Raster plots of active and inactive lever presses during a cue-associated sucrose seeking session (left), together with group mean data of total active (a) and inactive (i) lever presses during the 30 min seeking test. Food seeking was robust in control mice and did not differ in the 13 Hz mPFC group (effect of lever, $F_{1,12} = 163.7$, $P < 0.002$; group and lever \times group all non-significant; $n = 6/8$). Error bars, s.e.m.

Tip of fibers in:

- control mice • 13 Hz vHipp • 1 Hz vHipp
- 1 Hz mPFC • 13 Hz mPFC



Extended Data Figure 9 | Position of fibre optic cannula placements for mice used in tests of cue-seeking. Figure shows identified fibre optic cannula tip placements in controls (left) or mice that received different light protocols

targeting either vHipp or mPFC to NAc synapses (right). Note that when mice were used for electrophysiology recordings, fibre optic placements were visually confirmed but not recorded.

Extended Data Table 1 | ANOVA comparisons for (a) Figs 1–4 and (b) Fig. 5

a

Figure	Measure	ANOVA factors	Effect	P
1c	R/I	Cell (D1R-MSN, D2R-MSN) x Group (saline, cocaine)	Group	F(1,83)=10.86 0.001
			Cell	F(1,83)=3.18 <0.1
			Group x Cell	F(1,83)=6.08 <0.05
	A/N	Cell (D1R-MSN, D2R-MSN) x Group (saline, cocaine)	Group	F(1,83)=5.91 <0.05
			Cell	F(1,83)=11.91 0.001
			Group x Cell	F(1,83)=8.34 <0.01
2b	Amplitudes	Group (saline, cocaine) x Input (BLA, vHipp, mPFC)	Group	F(1,206)=6.44 <0.05
			Input	F(2,206)=18.01 <0.001
			Group x Input	F(2,206)=4.14 <0.05
2c	R/I	Group (saline, cocaine) x Input (BLA, vHipp, mPFC)	Group	F(1,56)=14.56 <0.001
			Input	F(2,56)=9.87 <0.001
			Group x Input	F(2,56)=18.03 <0.001
2d	A/N	Group (saline, cocaine) x Input (BLA, vHipp, mPFC)	Group	F(1,56)=1.64 n.s.
			Input	F(2,56)=11.44 <0.001
			Group x Input	F(2,56)=33.41 <0.001
3d	Amplitude @ +40 mV (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,45)=0.32 n.s.
			Protocol	F(2,45)=0.37 n.s.
			Group x Protocol	F(2,45)=0.23 n.s.
	RI (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,45)=1.39 n.s.
			Protocol	F(2,45)=0.64 n.s.
			Group x Protocol	F(2,45)=0.78 n.s.
3e	A/N (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,46)=59.43 <0.001
			Protocol	F(2,46)=9.76 <0.001
			Group x Protocol	F(2,46)=5.73 <0.01
3i	Amplitude @ +40 mV (mPFC)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,54)=15.72 <0.001
			Protocol	F(2,54)=3.5 <0.05
			Group x Protocol	F(2,54)=2.57 <0.1
	RI (mPFC)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,54)=33.23 <0.001
			Protocol	F(2,54)=13.18 <0.001
			Group x Protocol	F(2,54)=10.72 <0.001
3j	A/N (mPFC)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,54)=3.08 <0.1
			Protocol	F(2,54)=5.6 <0.01
			Group x Protocol	F(2,54)=6.0 <0.01
4b	Amplitude @ +40 mV (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,80)=9.92 <0.05
			Protocol	F(2,80)=3.37 <0.05
			Group x Protocol	F(2,80)=3.15 <0.05
	RI (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,80)=11.8 0.001
4c	A/N (vHipp)	Group (saline, cocaine) x Protocol (Control, 1 Hz, 13Hz)	Group	F(1,80)=13.5 <0.001
			Protocol	F(2,80)=5.6 <0.01
			Group x Protocol	F(2,80)=2.07 n.s.
4b	Amplitude @ +40 mV (mPFC)	Group (saline, cocaine) x Protocol (control, 1 Hz, 13 Hz)	Group	F(1,94)=12.1 0.001
			Protocol	F(2,94)=0.6 n.s.
			Group x Protocol	F(2,94)=3.96 <0.05
	RI (mPFC)	Group (saline, cocaine) x Protocol (control, 1 Hz, 13 Hz)	Group	F(1,94)=16.95 <0.001
			Protocol	F(2,94)=2.3 0.1
			Group x Protocol	F(2,94)=4.83 0.01
4c	A/N (mPFC)	Group (saline, cocaine) x Protocol (control, 1 Hz, 13 Hz)	Group	F(1,94)=9.12 <0.01
			Protocol	F(2,94)=2.2 n.s.
			Group x Protocol	F(2,94)=2.65 <0.1

b

Figure	Measure	Comparisons (ANOVA or t-test)	Effect	P
5b	Cocaine control vs. Saline control	Group (saline, cocaine) x Lever (active x inactive)	Lever	F(1,53)=21.7 <0.001
			Group	F(1,53)=16.8 <0.001
			Lever x Group	F(1,53)=22.9 <0.001
	Saline control	active vs. inactive	t(10)=0.43	n.s.
	Cocaine control	active vs. inactive	t(43)=9.5	<0.001
	Cocaine control vs. Saline control	active lever comparison	t(53)=4.6	<0.001
Cocaine control vs. Saline control	inactive lever comparison		t(53)=1.54	n.s.
	1 Hz vHipp vs. cocaine control	Lever (active, inactive) x Protocol (Control, Laser)	Lever	F(1,53)=45.7 <0.001
			Protocol	F(1,53)=4.7 <0.05
			Lever x Protocol	F(1,53)=5.13 <0.05
1 Hz vHipp	active vs. inactive		t(10)=3.6	<0.01
	1 Hz vHipp vs. cocaine control	active lever comparison	t(53)=2.3	<0.05
1 Hz vHipp vs. cocaine control	inactive lever comparison		t(53)=1.1	n.s.
	13 Hz vHipp vs. cocaine control	Lever (active, inactive) x Protocol (Control, Laser)	Lever	F(1,51)=30.8 <0.001
			Protocol	F(1,51)=4.12 <0.05
			Lever x protocol	F(1,51)=6.41 <0.05
13 Hz vHipp	active vs. inactive		t(8)=1.99	n.s.
	13 Hz vHipp vs. cocaine control	active lever comparison	t(51)=0.39	n.s.
13 Hz vHipp vs. cocaine control	inactive lever comparison		t(51)=4.48	<0.001
	1 Hz mPFC vs. cocaine control	Lever (active, inactive) x Protocol (Control, Laser)	Lever	F(1,53)=52.08 <0.001
			Protocol	F(1,53)=1.1 n.s.
			Lever x protocol	F(1,53)=3.08 n.s.
1 Hz mPFC	active vs. inactive		t(10)=4.16	0.002
	13 Hz mPFC vs. cocaine control	Lever (active, inactive) x Protocol (Control, Laser)	Lever	F(1,52)=23.2 <0.001
			Protocol	F(1,52)=16.7 <0.001
			Lever x protocol	F(1,52)=17.5 <0.001
13 Hz mPFC	active vs. inactive		t(9)=1.74	n.s.
	13 Hz mPFC vs. cocaine control	active lever comparison	t(52)=4.29	<0.001
13 Hz mPFC vs. cocaine control	inactive lever comparison		t(52)=2.19	<0.05
	5d	Lever (active, inactive) x Protocol (Control, Laser)	Lever	F(1,12)=163.7 <0.001
			Protocol	F(1,12)=0.022 n.s.
			Lever x protocol	F(1,12)=0.41 n.s.

Cell competition is a tumour suppressor mechanism in the thymus

Vera C. Martins^{1,2}, Katrin Busch¹, Dilafruz Juraeva³, Carmen Blum², Carolin Ludwig², Volker Rasche⁴, Felix Lasitschka⁵, Sergey E. Mastitsky³, Benedikt Brors³, Thomas Hielscher⁶, Hans Joerg Fehling² & Hans-Reimer Rodewald¹

Cell competition is an emerging principle underlying selection for cellular fitness during development and disease. Competition may be relevant for cancer, but an experimental link between defects in competition and tumorigenesis is elusive. In the thymus, T lymphocytes develop from precursors that are constantly replaced by bone-marrow-derived progenitors. Here we show that in mice this turnover is regulated by natural cell competition between ‘young’ bone-marrow-derived and ‘old’ thymus-resident progenitors that, although genetically identical, execute differential gene expression programs. Disruption of cell competition leads to progenitor self-renewal, upregulation of *Hmgal*, transformation, and T-cell acute lymphoblastic leukaemia (T-ALL) resembling the human disease in pathology, genomic lesions, leukaemia-associated transcripts, and activating mutations in *Notch1*. Hence, cell competition is a tumour suppressor mechanism in the thymus. Failure to select fit progenitors through cell competition may explain leukaemia in X-linked severe combined immune deficiency patients who showed thymus-autonomous T-cell development after therapy with gene-corrected autologous progenitors.

Cell competition, a process favouring fit over weak or harmful cells, was originally identified in *Drosophila*^{1–3}. Competition has been studied in tissue mosaics in which the interaction between genetically distinct cells differing in gene dosage eliminated one population (‘losers’) due to the presence of another (‘winners’). In cancer, following initial mutagenesis and transformation, competition may be involved in tumour cell selection, and progression^{2,3}.

Studies in haematopoietic stem and progenitor cells provided evidence for p53-mediated cell competition^{4,5} and for the importance of the ‘fitness’ of normal cells in preventing leukaemia from oncogene-bearing cells⁶. Because competition could provide a powerful mechanism of tissue surveillance, ‘natural’ cell competition (between genetically identical cells) might be a tumour-suppressor mechanism ensuring normal cell turnover, and preventing accumulation of mutations in ‘unopposed’ cells. Lack of competition may favour the formation rather than mere selection of faulty cells, but examples for natural cell competition, or consequences of its disruption, are elusive.

Thymus function normally depends on a continuous supply of T-cell progenitors migrating from the bone marrow to the thymus⁷. However, when deprived of progenitors from the bone marrow, the thymus can maintain autonomous T-cell development for several months^{8,9}. Hence, thymocyte turnover is regulated by bone marrow progenitor colonization. Acute lymphoblastic leukaemias (ALL) are the most frequent childhood malignancies. T-lineage ALL (T-ALL) arises by malignant transformation of T-cell progenitors¹⁰. Widespread genomic lesions in T-ALL reflect genomic instability in T-cell progenitors¹¹, and often affect genes regulating T-cell development¹². Notably, human T-ALL frequently bear somatic gain-of-function mutations in *NOTCH1*^{13,14}. Yet, clues on what goes astray when a normal thymus gives rise to leukaemia are lacking.

We report here that thymocyte turnover is regulated by natural cell competition. When relieved from outside competition, intrathymic precursors persist, self-renew, and eventually leukaemia emerges that resembles human T-ALL. The system links impaired cell competition to tumour formation, and enables molecular analyses of the progression from normal T-cell development via pre-malignant stages to leukaemia.

Progenitor deprivation causes T-ALL

Thymus autonomy is induced by exposing a normal thymus to an environment deficient for T-cell progenitors^{8,9} (Fig. 1a). Consequences of sustained progenitor deprivation (Fig. 1a) were analysed in thymus grafts, endogenous thymi and host spleens (Fig. 1b). T-cell development persisted (CD4⁺CD8⁺ double-positive (DP) thymocytes present), or was exhausted (recirculating CD4⁺ and CD8⁺ single-positive T cells present). CD4^{lo/+}CD8⁺ cells in the graft with large forward side scatter (not shown) suggested ‘pre-T-ALL’. Dissemination of these cells throughout lymphoid organs implied T-ALL development (Fig. 1b). The cells were mostly but not exclusively (Extended Data Fig. 1a) CD4^{lo/+}CD8⁺ and had blast morphology (Fig. 1c). Mice developed tumours at the site of grafting (Extended Data Fig. 1b), splenomegaly (Fig. 1d), infiltrations into bone marrow and solid organs as seen by anaemic bone cavities (Fig. 1e), pale livers (Fig. 1f) and kidneys (not shown). Spleen volumes and cellularity were increased about tenfold compared with normal spleens (Extended Data Fig. 1c, d). Leukaemic cells proliferated (Extended Data Fig. 1e), and were broadly localized in the spleen (Extended Data Fig. 1f). The liver showed perivascular and periportal infiltrates of CD8⁺ neoplastic cells (Extended Data Fig. 1g, h). Primary T-ALL samples were transplantable into nude mice (Extended Data Figs 1i and 2). Because T-cell antigen receptor (TCR) β loci had one (12/20 T-ALL samples) or two (8/20 samples) rearrangements (Extended Data Fig. 3), T-ALL originated from one or two clones at most. In 13/20 (65%) samples, we found at least one productive TCR- β rearrangement, suggesting that transformation can occur before or after pre-TCR selection¹⁵.

Bone marrow progenitor fitness and T-ALL development in the thymus

We characterized the impact of the relative fitness of bone-marrow-derived progenitors on leukaemia development. Whereas progenitors in the thymus were always wild type, the fitness of bone marrow progenitors increased in the following order: $Rag2^{-/-}\gamma_c^{-/-}Kit^{W/W^v} < Rag2^{-/-}\gamma_c^{-/-} < \gamma_c^{-/-} < Kit^{W/W^v} < Rag1^{-/-}$. These mutants are deficient

¹Division of Cellular Immunology, German Cancer Research Center, D-69120 Heidelberg, Germany. ²Institute of Immunology, University of Ulm, D-89081 Ulm, Germany. ³Division of Theoretical Bioinformatics, German Cancer Research Center, D-69120 Heidelberg, Germany. ⁴Core Facility Small Animal MRI, University of Ulm, D-89081 Ulm, Germany. ⁵Institute of Pathology, University Hospital Heidelberg, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany. ⁶Division of Biostatistics, German Cancer Research Center, D-69120 Heidelberg, Germany.

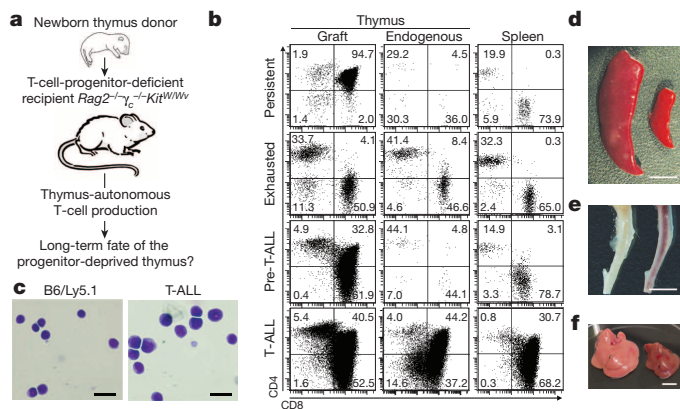


Figure 1 | Progenitor deprivation causes T-ALL. **a**, Experimental design. **b**, Cell phenotypes, pre-gated for thymus donor origin, in thymus grafts, endogenous thymi and spleens of recipient mice. Numbers show percentages of cells in the quadrants. Data are representative for one out of two experiments analysing a total of 11 mice 16 to 18 weeks after transplantation. **c**, Cytopins of control (B6/Ly5.1) (left) and the T-ALL⁺ spleens (right) (scale bar, 20 μ m). **d–f**, Spleens (**d**), tibias (**e**) and livers (**f**) from a control mouse (B6/Ly5.1) (right), and a T-ALL⁺ recipient (left). Data shown (**d–f**) are representative for all T-ALL⁺ mice. Scale bars, 0.5 cm in **d**, **e**; 1 cm in **f**.

in responding to essential thymus cytokines, Kit ligand and interleukin 7 (IL-7)¹⁶, and/or have a developmental block caused by recombination activating gene (*Rag1* or *Rag2*) ablation. In the least competitive environment, 64% of all recipients developed T-ALL (Fig. 2a). Incidences were 50% in *Rag2*^{-/-} γ_c ^{-/-}, and 38% in γ_c ^{-/-} (mouse model of human X-linked severe combined immune deficiency (SCID)) hosts. Because the IL-7 receptor (IL-7r) is composed of IL-7 α plus γ_c chains, we tested *Il7r α* ^{-/-} hosts, which were unable to suppress leukaemia (Fig. 2a). Hence, the incidence of leukaemia was determined by the failure of bone-marrow-derived progenitors to outcompete intrathymic progenitors by IL-7 responsiveness. T-ALL did not develop in *Rag1*^{-/-} or *Kit*^{W/W^v} mutants, in which the bone marrow fully replaces intrathymic progenitors⁸, and none of the thymus donor strains were 'prone' for T-ALL (Fig. 2a) (Supplementary Discussion).

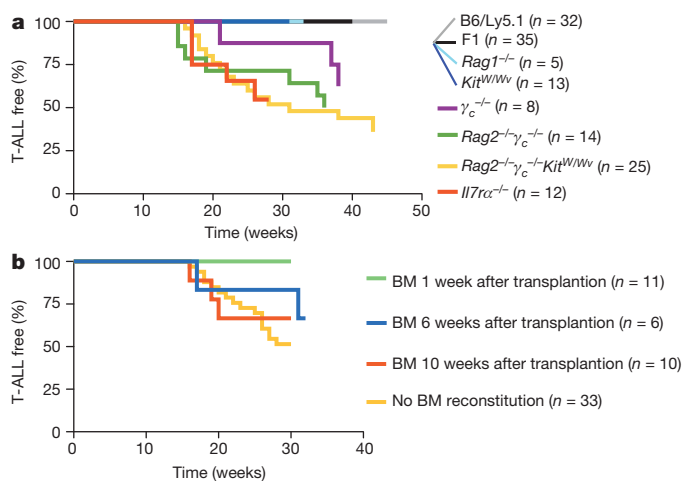


Figure 2 | T-ALL emergence and its prevention by bone marrow reconstitution. **a**, Kaplan–Meier plot of T-ALL incidence after grafting of wild-type thymi (as in Fig. 1a) into indicated recipients, and B6/Ly5.1 and F1(WB-*Kit*^{+/+} \times B6/Ly5.1) controls (*n* are indicated). **b**, Three groups of *Rag2*^{-/-} γ_c ^{-/-} *Kit*^{W/W^v} mice received first a wild-type thymus transplant, and 1, 6, or 10 weeks later were reconstituted with wild-type bone marrow (BM) cells (Extended Data Fig. 4). The groups were followed for up to 32 weeks. The yellow curve summarizes the data for the observation period of 30 weeks for all analysed *Rag2*^{-/-} γ_c ^{-/-} *Kit*^{W/W^v} not receiving bone marrow after thymus transplantation.

If lack of replacement of intrathymic progenitors by competitive progenitors from the bone marrow was responsible for T-ALL development, restoration with competitive bone marrow progenitors should prevent T-ALL. We reconstituted thymus-grafted *Rag2*^{-/-} γ_c ^{-/-} *Kit*^{W/W^v} mice (Fig. 1a) with wild-type bone marrow at different time points (1, 6, or 10 weeks) after thymus grafting (Extended Data Fig. 4a). These recipients accept bone marrow without prior myeloablation¹⁷, avoiding effects that irradiation itself would have on leukaemia development. Only mice reconstituted with bone marrow 1 week after thymus transplantation were protected from T-ALL (Fig. 2b). Collectively, disease could be prevented by early restoration of normal competitive bone marrow, but by 6 weeks after onset of thymus autonomy transformation became irreversible (Extended Data Fig. 4b).

Molecular characterization of T-ALL

To monitor the emergence of T-ALL from the thymus, we transplanted reporter thymus harbouring a green fluorescent protein (GFP) transgene driven by the *Rag2* promoter (RAG2p-GFP)¹⁸, which allows visualization of thymic emigrants. At 17 weeks, GFP⁺ CD4^{lo} CD8⁺ cells appeared in the blood (Extended Data Fig. 5a). One week later, this recipient showed clinical signs of disease, and leukaemic cells had disseminated into the bone marrow and spleen (Extended Data Fig. 5b). Based on this reporter, and on reverse transcription followed by PCR (RT-PCR) analysis (not shown), T-ALL continued to express *Rag2* outside the thymus, and after transfer into nude recipients (Extended Data Fig. 2a).

To analyse the impact of *Rag* expression, which has been linked to oncogenic translocations¹⁹, on leukaemia incidence (Fig. 3a, b) and on genomic lesions (Fig. 3c, d), we grafted *Rag1*^{-/-} TCR transgenic *tg(OT-I)* thymi into *Rag2*^{-/-} γ_c ^{-/-} *Kit*^{W/W^v} mice. On a *Rag*-deficient background, the transgenic TCR promotes differentiation beyond CD3⁺ CD4⁺ CD8⁺ triple-negative (TN) stages. T-ALL development in *Rag1*^{-/-} *tg(OT-I)* thymus grafts (Fig. 3a, b) excluded *Rag* expression as a prerequisite for leukaemia. Developmental progression beyond TN stages, a scenario consistent with T-ALL phenotypes (Fig. 1, Extended Data Fig. 1a), was also not required for transformation because *Rag1*^{-/-} progenitors lacking the transgenic TCR developed T-ALL (Fig. 3a, b). Collectively, T-ALL had an immature CD8 single-positive (ISP)/DP-like surface phenotype, but transformation did not require progression beyond TN stages.

In human and mouse T-ALL, genomic gains and losses reflect genomic instability^{19–23}. Indeed, array comparative genomic hybridization (aCGH) of T-ALL revealed lesions (Fig. 3c) in genes known to be involved in cancer (Supplementary Tables 1 and 2). Hence, progenitors, once relieved from cell competition, become genomically unstable. *Rag1* expression contributed to illegitimate or enhanced DNA recombination because lesion numbers were lower in *Rag1*^{-/-} T-ALL (Fig. 3d).

Typical for T-ALL, gain-of-function mutations in *Notch1* affect the heterodimerization (HD), transactivation (TAD) and PEST domains^{14,24,25} (Fig. 3e), resulting in ligand-independent signalling, or greater intracellular protein stability^{26,27}. Indeed, we found at least one heterozygous mutation in 22/28 T-ALL (triangles in Fig. 3e). Affected domains, DNA mutation types and putative functional consequences (Extended Data Fig. 6a) were similar to mutations in human T-ALL¹⁴. In murine models, 5' genomic deletions in *Notch1* drive alternative transcripts, resulting in active intracellular *Notch1*^{28–30}. Such 5' deletions were detected in 13/25 *Rag1*^{+/+} mice but in none of the 5 *Rag1*^{-/-} T-ALL (Extended Data Fig. 6b, c). *Notch1* was expressed from a 5' alternative promoter³⁰ in T-ALL but not in normal thymus (Extended Data Fig. 6d). Collectively, 79% of T-ALL showed gain-of-function mutations (*Rag1*-independent), and 52% had 5' deletions (*Rag1*-dependent) in the *Notch1* gene (Fig. 3f).

Intrathymic origin of T-ALL and self-renewal

Leukaemia cells are considered developmentally 'frozen', and their phenotype is thought to reflect their stage of origin. We used global gene expression profiling to compare T-ALL samples (*n* = 10) with cells from different stages of T-cell development, ranging from early thymic

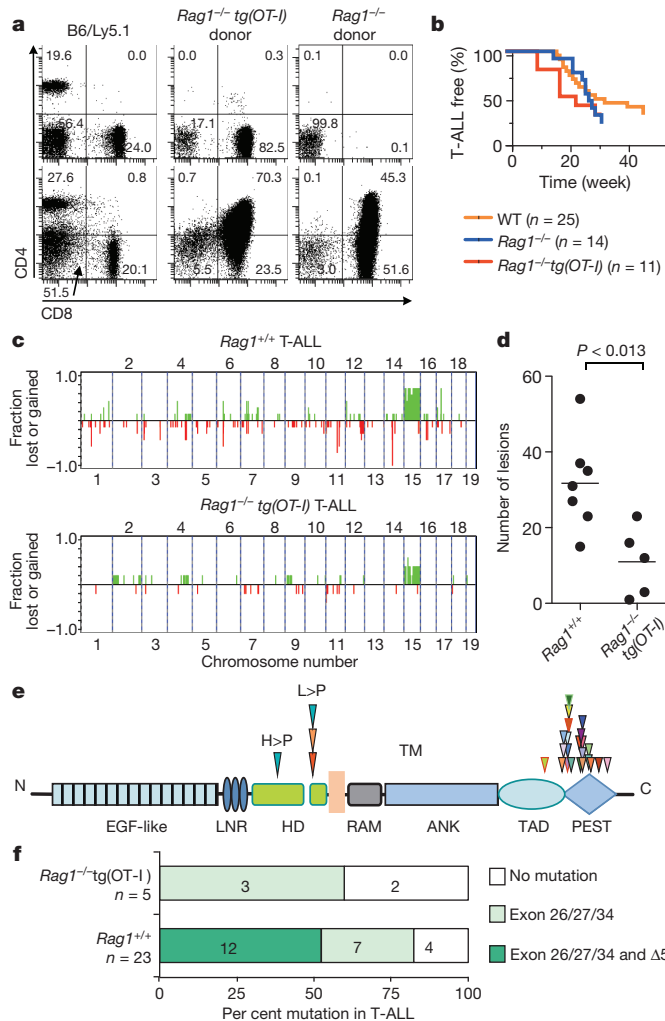


Figure 3 | Molecular characterization of T-ALL. **a**, Thymi from newborn *Rag1*^{-/-} tg(OT-I) or *Rag1*^{-/-} donors were transplanted into *Rag2*^{-/-} γ_c ^{-/-} Kit^{W/W^v} mice (as in Fig. 1a). 8 weeks later, peripheral blood was analysed for T cells (upper row). At the onset of T-ALL, spleens were analysed (lower row). B6/Ly5.1 mice served as staining controls. Data are representative for T-ALL⁺ mice indicated. **b**, Kaplan-Meier plots of T-ALL development in *Rag2*^{-/-} γ_c ^{-/-} Kit^{W/W^v} recipients of wild-type, *Rag1*^{-/-} tg(OT-I), or *Rag1*^{-/-} thymus grafts. None of the curves in **b** were significantly different. **c**, *Rag1*^{+/+} (n = 7) and *Rag1*^{-/-} tg(OT-I) (n = 5) T-ALL samples were analysed by aCGH. Fractions of mice with gains (green) and losses (red) are shown by chromosomal positions. **d**, Number of lesions in T-ALL, excluding trisomy 15, were significantly higher in *Rag1*^{+/+} compared to *Rag1*^{-/-} tg(OT-I) T-ALL (Mann-Whitney-Wilcoxon test). **e**, HD, TAD and PEST domains of Notch1 are depicted. Each triangle corresponds to one mutation. Each colour represents a single T-ALL (see Extended Data Fig. 6a). **f**, Frequencies of *Notch1* mutations (exon 26/27/35; triangles in e), and deletions ($\Delta 5'$) (see Extended Data Fig. 6b, c) in T-ALL.

progenitor (ETP), to TN CD4⁺ CD25⁺ (TN2), to CD44⁺ CD25⁺ (TN3), to CD44⁺ CD25⁻ (TN4), to ISP, to DP cells. Principle component analysis (PCA) revealed the closest relationship of T-ALL to TN3/TN4/ISP stages (Fig. 4a), indicating that upon progenitor deprivation stages between TN3 to ISP enter self-renewal, and eventually undergo malignant transformation. Self-renewal was analysed by intrathymic injection of TN cells (containing TN1–4), or more mature DP/SP positive cells, each sorted from 10-week-old autonomous thymus grafts (Fig. 4b). Whereas DP/SP donor cells retained their phenotype with no evidence for further development, and limited expansion *in vivo* (~ fivefold), TN donor cells gave rise to the major downstream stages (DP and SP), and their expansion was ~ 500-fold. The data imply that the self-renewing cells that maintain the residual thymus cellularity and T-cell development

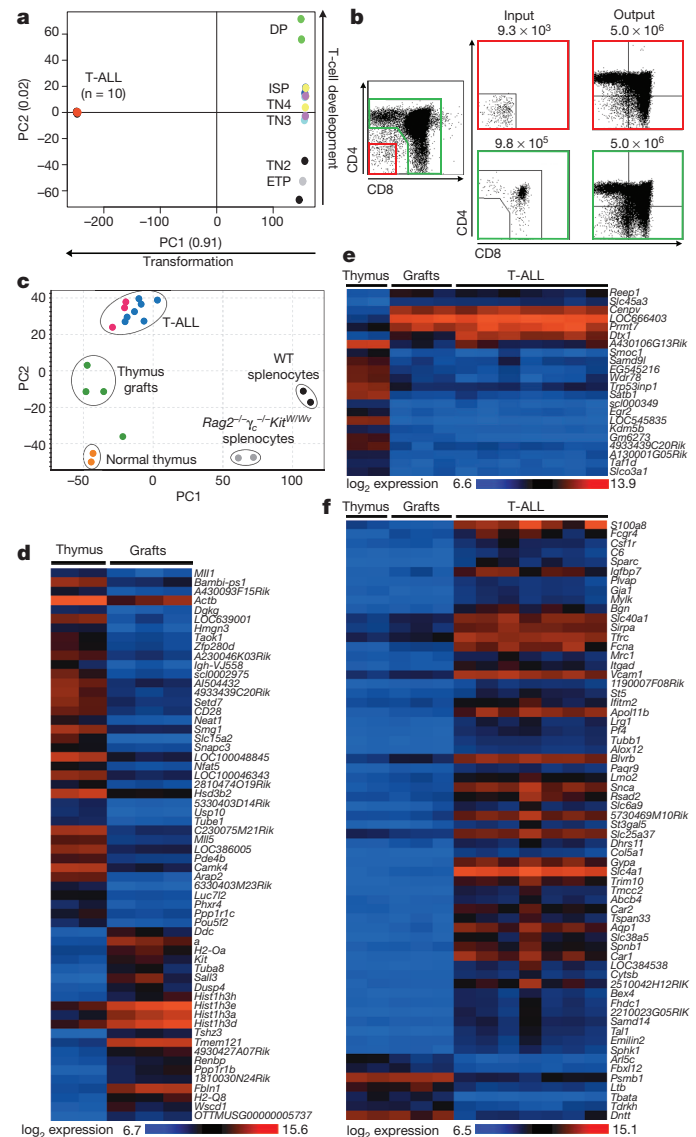


Figure 4 | Origin of T-ALL, self-renewal in the thymus, and transcriptome analyses during disease progression. **a**, PCA for global gene expression of T-ALL samples (n = 10; appearing as one dot), and stages of T-cell development, indicated by unique colours for each population (n = 2) (see text). PC1 reveals the highest variance between leukaemic samples and thymocyte subpopulations. PC2 reflects progression along T-cell development, revealing the highest association of T-ALL to TN3/TN4/ISP stages. **b**, CD3⁺ CD4⁺ CD8⁻ TN, or CD4⁺ CD8⁺ DP together with CD4⁺ CD8⁻ and CD4⁺ CD8⁺ SP populations were sorted from 10-week progenitor-deprived thymi (n = 7) (left), reanalysed (middle), and injected intrathymically into irradiated recipients. Phenotypes of progeny from TN-injected (n = 3), or DP/SP-injected (n = 2) recipients were analysed on day 21 (right). Numbers above the panels are cell numbers injected (input), and retrieved (output) per mouse. **c**, PCA of transcriptomes revealed segregation by organ (thymus, spleen) and by condition (normal thymus, progenitor-deprived thymus grafts, T-ALL). *Rag1*^{+/+} or *Rag1*^{-/-} T-ALL are depicted in blue or red dots. **d–f**, Three pairwise comparisons were overlaid in a Venn diagram (Extended Data Fig. 7), revealing unique and common transcriptional changes. The heat maps show (threshold \geq fourfold) changes unique for the transition from the normal thymus (n = 2) to thymus grafts (n = 3) (**d**), changes maintained into the T-ALL stage (n = 7) (**e**), and changes exclusive for T-ALL (**f**).

are contained in the TN compartment. It is difficult to distinguish self-renewal in non-transformed T-cell progenitors from pre-malignant behaviour of cells with retained differentiation potential. Nevertheless, these data are consistent with self-renewal in the TN compartment from which T-ALL arise (Fig. 4a).

Pre-malignant and leukaemic gene expression

PCA of gene expression in normal thymus with regular progenitor turnover, in thymus devoid of extrinsic progenitor competition for 10 weeks, and in fully malignant T-ALL (Fig. 4c–f) revealed sample clustering according to experimental groups, demonstrating consistent gene expression changes during deprivation and subsequent transformation (Fig. 4c). Three pairwise comparisons (see Extended Data Fig. 7 for rationale of analysis) revealed transcriptional changes (threshold \geq fourfold) unique for the transition from the normal to the progenitor-deprived thymus (Fig. 4d), or maintained into the T-ALL stage (Fig. 4e). Changes exclusive for T-ALL (Fig. 4f) included many cancer- and leukaemia-associated genes (see Supplementary Discussion), notably *Lmo2* and *Tal1* (Fig. 4f). Lowering the threshold to \geq two-fold differential expression, and focusing on cancer-associated genes, uncovered additional information (Supplementary Tables 3–5). *Ikaros* (also known as *Irfz1*), *Pten* and *Rb1* expression was reduced already in thymus grafts (Supplementary Table 3), whereas *Lyl1* and *Myc* expression increased only in T-ALL (Supplementary Table 5). In summary, the gene expression profile in T-ALL was, in part, already initiated in the thymus upon progenitor deprivation.

Gene expression data (Fig. 4f) were used to correlate this murine T-ALL to subgroups of human T-ALL³¹. Spearman rank correlation coefficients revealed significant correlations between our T-ALL and all human T-ALL subtypes, but not to human B-ALL³² (Extended Data Fig. 8a, b), excluding the possibility that correlations between murine and human T-ALL reflect a ‘transformed state’. Important oncogenes in human T-ALL include *TAL1*, *LMO2*, *TLX1* and *LYL1*. *Tal1* and *Lmo2* were strongly upregulated in the murine T-ALL, *Lyl1* was also significantly but more weakly expressed whereas *tlx1* was not expressed (Extended Data Fig. 8c). Hence, similar oncogenic pathways were found in T-ALL arising in the bone-marrow-progenitor-deprived thymus compared to

human subgroups of T-ALL, but not B-ALL. The homologues of the most commonly found oncogenes in human T-ALL, *Tal1* and *Lmo2*, were strongly induced in the murine T-ALL.

Hallmarks of cell competition driving thymocyte turnover

Thymus autonomy^{8,9} and subsequent T-ALL development occur when bone marrow progenitors are unresponsive to IL-7, whereas intrathymic progenitors receive IL-7 receptor signals (Fig. 2). We constructed mixed chimaeras composed of titrated numbers of wild-type ($\gamma_c^{+/+}$) and IL-7r-deficient ($\gamma_c^{-/-}$) donor bone marrow, which revealed the sensitivity and stage (TN2/TN3) of IL-7-driven selection of responsive over unresponsive precursors (Extended Data Fig. 9).

Kinetic analysis of the physiological replacement of thymus-resident by bone-marrow-derived progenitors (both wild type) revealed day 21 as the ‘turning point’ for total thymocytes (Fig. 5a), and day 14 for TN2/TN3 cells (Fig. 5e). The turnover was completed by day 35 (Fig. 5b, f). By comparison, when intrathymic progenitors (wild type) lacked competition from bone marrow progenitors (*Rag2*^{-/-} $\gamma_c^{-/-}$ *Kit*^{W/W^v hosts), both total thymocytes (Fig. 5c) and TN2/TN3 cells (Fig. 5g) persisted beyond day 35. The fact that one cell population (thymus-resident progenitors) was lost in the presence, but not in the absence of another population (bone-marrow-derived progenitors), indicates that normal T-cell progenitor turnover is regulated by cell competition. The extent of the competition-dependent cell loss is illustrated by the shaded areas under the curves (Fig. 5d, h).}

Thymus-resident and bone-marrow-derived TN2/TN3 cells differ in their time of residency in the thymus (Fig. 5e), and therefore in the duration for which they have been engaged in T-cell development. We therefore operationally refer to thymus-resident TN2/TN3 cells as

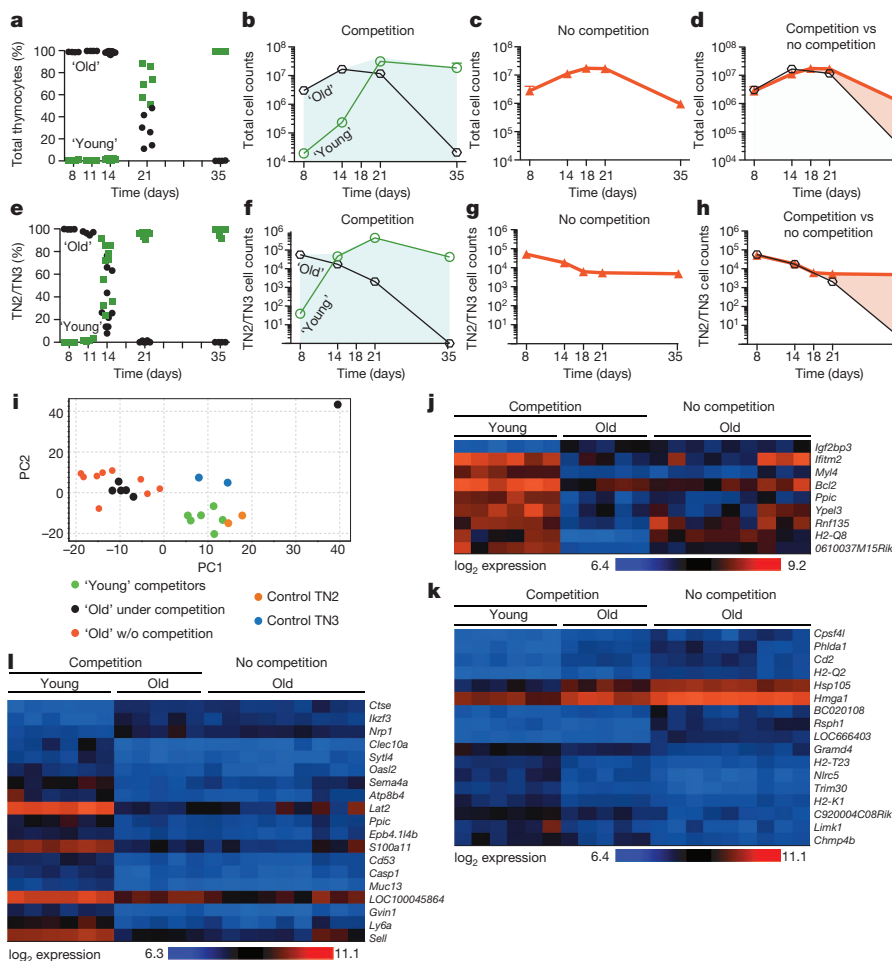


Figure 5 | Cell competition regulates thymocyte turnover. Wild-type thymi were transplanted into competitive (wild type) (a, b, e, f) or non-competitive (*Rag2*^{-/-} $\gamma_c^{-/-}$ *Kit*^{W/W^v) environments (c, g). a, b, e, f, Percentages (a, e) and cell counts (b, f) of old, graft-derived (black line) versus young, bone-marrow-derived (green line) total thymocytes (a, b) or TN2/TN3 progenitors (e, f) over time. Shaded areas represent total thymocytes (b) or TN2/TN3 progenitors (f), regardless of origin. c, g, Number of total thymocytes (c) or TN2/TN3 progenitors (g) in the absence of competition. d, h, Numbers of graft-derived thymocytes (d) or graft-derived TN2/TN3 cells (h) in the presence (black line) or absence (red line) of competition. Shaded area highlights the effect of cell competition. i, PCA of TN2/TN3 that were young, old under competition, or old without competition, and control TN2 and TN3. j–l, Pairwise comparisons between young ($n = 6$) and old cells in the presence ($n = 5$) or absence ($n = 9$) of cell competition (threshold \geq twofold) were overlaid in a Venn diagram (Extended Data Fig. 10a). Including additionally the comparison of control TN2 versus TN3 cells allowed exclusion of changes associated with the TN2-to-TN3 transition. This yielded lists of competition-responsive genes (j), genes specific for old cells in the absence of competition (k) and genes associated with old cells regardless of competition (l).}

'old', and to bone-marrow-derived TN2/TN3 cells as 'young'. We analysed global gene expression for old TN2/TN3 cells that were under competition from young TN2/TN3 cells (as in Fig. 5f) or not (as in Fig. 5g), and for young TN2/TN3 cells. PCA revealed distinct clusters for young and for old TN2/TN3, whereas the clusters for old TN2/TN3 with or without competition overlapped (Fig. 5i). Pairwise comparisons (Extended Data Fig. 10a) revealed the impact of competition on old TN2/TN3 cells. Using a threshold of \geq twofold, we identified 9 genes with specifically altered expression in old TN2/TN3 cells under competition (Fig. 5j). Interestingly, the anti-apoptotic gene *Bcl2* was among these genes, and this translated into reduced expression of the Bcl2 protein. Consistently, frequencies of annexin⁺ cells were increased in old compared to young TN2/TN3 cells (Extended Data Fig. 10b–d). Because Bcl2 expression is downstream from IL-7r³³, these data suggest a link between the natural replacement process and the observed effects of IL-7r signalling on competition^{8,9}. The functional implications of other identified genes, provisionally referred to as 'competition-regulated genes' (Fig. 5j), will require further studies.

Changes specific for old TN2/TN3 cells in the absence of competition were also identified (Fig. 5k). Remarkably, the most upregulated gene was high-mobility group A1 (*Hmgal*), a transcriptional master regulator overexpressed in stem cells and in many cancer types, including human T-ALL^{34,35}. Finally, we also identified 19 genes that were differentially expressed comparing young and old T-cell progenitors, regardless of competition (Fig. 5l).

Discussion

Our findings identify cell competition as a mechanism driving the physiological substitution of thymus-resident progenitors ('old') by bone-marrow-derived ('young') progenitors. We have characterized this dynamic process in kinetic and quantitative terms. The results demonstrate that old and young progenitors, although genetically identical and phenotypically positioned at the same developmental stage, differ selectively in gene expression, which may reflect their distinct past and current 'cellular experiences', such as dwell times in the thymus, cumulative proliferation numbers, cellular or genetic insults suffered during *V(D)J* recombination and DNA repair, growth factor dependency, stromal niche support, and others. Levels of Bcl2 were reduced in old compared to young progenitors, which was consistent with higher frequencies of apoptotic cells in the old compared to the young. Because Bcl2 expression is downstream of IL-7r signalling, lower Bcl2 expression suggests that IL-7 availability is limited for, or less utilized by, old progenitors. This provides an attractive link between the experiments in which IL-7-unresponsive bone marrow progenitors could not inhibit thymus autonomy^{8,9} and natural cell competition, in which both thymus and bone marrow progenitors are wild type. We took advantage of mice lacking T-cell progenitors in the bone marrow to relieve old wild-type progenitors in the thymus from competition. In such 'unopposed' old progenitors, we identified several genes that partially resemble the 'young state' (Fig. 5j), and others that distinguish specifically old progenitors with competition from those without (Fig. 5k). Hence, old progenitors seem to sense and react to competition by young progenitors.

The consequences of disrupted cell competition on intrathymic T-cell progenitors are of particular interest. Already after 14 days without competition, old cells strongly upregulate *Hmgal*, a key regulator of stemness that is highly expressed in many cancers. *Hmgal* encodes non-histone chromatin proteins that act as architectural transcriptional factors. These not only regulate diverse developmental processes but also orchestrate malignant transformation. *Hmgal* transgenic mice develop aggressive T-ALL that is accelerated by loss of *Cdkn2a*^{34,35}, one of the loci affected in our model (Supplementary Table 2). At the cellular level, progenitor deprivation enables self-renewal at TN stages, as shown by TN2/TN3 persistence in the absence of competition, and the fact that even after 10 weeks without new progenitor import TN cells gave rise to DP and SP stages upon adoptive transfer. By this time, many genes are deregulated compared to a normal thymus, with some changes carried

over into T-ALL. On the basis of clinical, pathological, cellular and molecular criteria, T-ALL developing in mice by disruption of cell competition in the thymus share many properties with human T-ALL. Collectively, we propose that cell competition protects the T-cell progenitor population from *in situ* ageing, and in this manner is a tumour-suppressor mechanism.

Promotion of leukaemia by thymus autonomy could explain T-ALL in X-linked SCID patients in the wake of gene therapy. X-linked SCID patients carry germline mutations in γ_c , blocking early T-cell development in the human thymus. Following transplantation of autologous, gene-corrected stem or progenitor cells without myeloablation, all T cells, but no myeloid or B cells, were derived from γ_c^+ transplanted cells³⁶. Vector integration near *Lmo2* has been held responsible for T-ALL that developed in 5/20 treated patients^{22,37}, but leukaemias also harboured widespread genomic abnormalities unrelated to the viral integration^{22,23}. Given the sustained T-cell development in the patients' thymus, with no signs of bone marrow engraftment, and no evidence for T-cell progenitor replenishment, γ_c^+ progenitors probably only colonized the thymus where they prevailed long-term over γ_c^- (endogenous) progenitors. Our transplantation of wild-type thymus (containing competent γ_c^+ progenitors) into $\gamma_c^{-/-}$ mice closely reflects this situation. In contrast, none of the patients receiving gene therapy to correct adenosine deaminase (ADA) deficiency developed leukaemia, and all were pre-conditioned to permit HSC engraftment. We conclude that lack of progenitor supply from the bone marrow, combined with autonomous T-cell development in the thymus, favours progenitor transformation. Similarly, mutations in mice associated with development of leukaemia, notably *Lmo2* overexpression³⁸ or loss of *Tcf1* (also known as *Hnf1a*)³⁹, may force T-cell progenitors into self-renewal and enhance their competitiveness. This enhanced competitiveness then prevents normal cell turnover and promotes transformation. Altogether, the findings reported here for the thymus link defects in cell competition to tumorigenesis. It might be interesting to apply this framework to cancers arising in other cells or tissues maintained by stem and progenitor reservoirs.

METHODS SUMMARY

All mouse experiments were approved by Regierungspräsidien Tübingen or Karlsruhe. Recipient mice were C57BL/6, *Rag2*^{-/-} $\gamma_c^{-/-}$ *Kit*^{W/W^v}, *Kit*^{W/W^v}, *Rag2*^{-/-} $\gamma_c^{-/-}$, $\gamma_c^{-/-}$, *Il7ra*^{-/-}, *Rag1*^{-/-}, *Rag1*^{-/-}*tg(OT-I)* (all CD45.2⁺). Thymus donors were B6/Ly5.1 (H-2^b; CD45.1⁺), F1(WB-Kit^{+/+} \times B6/Ly5.1), F1(WB-Kit^{+/+} \times B6/Ly5.1-RAG2p-GFP transgenic), or *Rag1*^{-/-}, or *Rag1*^{-/-}*tg(OT-I)* mice. Survival curves and statistical analyses were calculated with Prism 5. For transcriptome analysis RNA was isolated using RNeasy B (IsoTex Diagnostics). Labelled complementary RNA was hybridized to Illumina MouseWG-6 v2.0 BeadChips. Data were analysed with Chipster v2.9. For aCGH, labelled DNA was hybridized to Agilent Mouse Genome aCGH Microarrays. Analyses were carried out using R 2.15/Bioconductor 2.11 with add-on packages limma, aCGH, snapCGH, DNACopy and cghMCR.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 May 2013; accepted 10 April 2014.

Published online 14 May 2014.

- Johnston, L. A. Competitive interactions between cells: death, growth, and geography. *Science* **324**, 1679–1682 (2009).
- Baker, N. E. Cell competition. *Curr. Biol.* **21**, R11–R15 (2011).
- Lavayer, R. & Moreno, E. Mechanisms of cell competition: themes and variations. *J. Cell Biol.* **200**, 689–698 (2013).
- Marusyk, A., Porter, C. C., Zaberezhnyy, V. & DeGregori, J. Irradiation selects for p53-deficient hematopoietic progenitors. *PLoS Biol.* **8**, e1000324 (2010).
- Bondar, T. & Medzhitov, R. p53-mediated hematopoietic stem and progenitor cell competition. *Cell Stem Cell* **6**, 309–322 (2010).
- Porter, C. C., Baturin, D., Choudhary, R. & DeGregori, J. Relative fitness of hematopoietic progenitors influences leukemia progression. *Leukemia* **25**, 891–895 (2011).
- Frey, J. R., Ernst, B., Surh, C. D. & Sprent, J. Thymus-grafted SCID mice show transient thymopoiesis and limited depletion of V beta 11+ T cells. *J. Exp. Med.* **175**, 1067–1071 (1992).
- Martins, V. C. *et al.* Thymus-autonomous T cell development in the absence of progenitor import. *J. Exp. Med.* **209**, 1409–1417 (2012).

9. Peaudecerf, L. *et al.* Thymocytes may persist and differentiate without any input from bone marrow progenitors. *J. Exp. Med.* **209**, 1401–1408 (2012).
10. Pui, C. H., Robison, L. L. & Look, A. T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030–1043 (2008).
11. Van Vlierberghe, P. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).
12. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature Rev. Immunol.* **8**, 380–390 (2008).
13. Koch, U. & Radtke, F. Mechanisms of T cell development and transformation. *Annu. Rev. Cell Dev. Biol.* **27**, 539–562 (2011).
14. Weng, A. P. *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
15. von Boehmer, H. & Fehling, H. J. Structure and function of the pre-T cell receptor. *Annu. Rev. Immunol.* **15**, 433–452 (1997).
16. Rodewald, H. R., Ogawa, M., Haller, C., Waskow, C. & DiSanto, J. P. Pro-thymocyte expansion by c-kit and the common cytokine receptor gamma chain is essential for repertoire formation. *Immunity* **6**, 265–272 (1997).
17. Waskow, C. *et al.* Hematopoietic stem cell transplantation without irradiation. *Nature Methods* **6**, 267–269 (2009).
18. Yu, W. *et al.* Continued RAG expression in late stages of B cell development and no apparent re-induction after immunization. *Nature* **400**, 682–687 (1999).
19. Zhang, Y. *et al.* The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv. Immunol.* **106**, 93–133 (2010).
20. Maser, R. S. *et al.* Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447**, 966–971 (2007).
21. De Keersmaecker, K. *et al.* The *TLX1* oncogene drives aneuploidy in T cell transformation. *Nature Med.* **16**, 1321–1327 (2010).
22. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).
23. Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**, 3143–3150 (2008).
24. O'Neil, J. *et al.* Activating *Notch1* mutations in mouse models of T-ALL. *Blood* **107**, 781–785 (2006).
25. Lin, Y. W., Nichols, R. A., Letterio, J. J. & Aplan, P. D. *Notch1* mutations are important for leukemic transformation in murine models of precursor-T leukemia/lymphoma. *Blood* **107**, 2540–2543 (2006).
26. Grabher, C., von Boehmer, H. & Look, A. T. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nature Rev. Cancer* **6**, 347–359 (2006).
27. Aster, J. C., Blacklow, S. C. & Pear, W. S. Notch signalling in T-cell lymphoblastic leukaemia/lymphoma and other haematological malignancies. *J. Pathol.* **223**, 263–274 (2011).
28. Tsuji, H. *et al.* Rag-dependent and Rag-independent mechanisms of *Notch1* rearrangement in thymic lymphomas of *Atm*^{-/-} and scid mice. *Mutat. Res.* **660**, 22–32 (2009).
29. Ashworth, T. D. *et al.* Deletion-based mechanisms of Notch1 activation in T-ALL: key roles for RAG recombinase and a conserved internal translational start site in *Notch1*. *Blood* **116**, 5455–5464 (2010).
30. Gómez-del Arco, P. *et al.* Alternative promoter usage at the *Notch1* locus supports ligand-independent signaling in T cell development and leukemogenesis. *Immunity* **33**, 685–698 (2010).
31. Homminga, I. *et al.* Integrated transcript and genome analyses reveal *NKX2-1* and *MEF2C* as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**, 484–497 (2011).
32. Stam, R. W. *et al.* Gene expression profiling-based dissection of MLL translocated and MLL germline acute lymphoblastic leukemia in infants. *Blood* **115**, 2835–2844 (2010).
33. Jiang, Q. *et al.* Distinct regions of the interleukin-7 receptor regulate different Bcl2 family members. *Mol. Cell. Biol.* **24**, 6501–6513 (2004).
34. Di Cello, F. *et al.* Inactivation of the *Cdkn2a* locus cooperates with *HMGAI* to drive T-cell leukemogenesis. *Leuk. Lymphoma* **54**, 1762–1768 (2013).
35. Fusco, A. & Fedele, M. Roles of HMG proteins in cancer. *Nature Rev. Cancer* **7**, 899–910 (2007).
36. Cavazzana-Calvo, M. *et al.* Is normal hematopoiesis maintained solely by long-term multipotent stem cells? *Blood* **117**, 4420–4424 (2011).
37. Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
38. McCormack, M. P. *et al.* The *Lmo2* oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science* **327**, 879–883 (2010).
39. Tiemessen, M. M. *et al.* The nuclear effector of Wnt-signaling, Tcf1, functions as a T-cell-specific tumor suppressor for development of lymphomas. *PLoS Biol.* **10**, e1001430 (2012).
40. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucl. Acids Res.* **30**, 207–210 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Tietz and T. Arnsperger for technical assistance, T. Ashworth and J. Aster, M. Nussenzweig and P. Fink for materials and mice, the animal facilities for mouse husbandry, S. Henze, O. Heil and E. Korpelainen for expression array experiments and support on Chipster. We thank B. Edgar, L. Johnston, T. Boehm, A. Rudensky, R. Medzhitov, M. Muckenthaler, A. Kulodzki, T. Sanda, T. Look, P. Lichter, T. Feyerabend and I. Rode for help and discussions. H.-R.R. was supported by ERC Advanced Grant No. 233074, DFG-SFB 938-project L, and the Helmholtz PCCC Alliance.

Author Contributions V.C.M. designed and performed experiments, and wrote the paper; K.B., C.B. and C.L. performed experiments; V.C.M., D.J., S.E.M., B.B. and T.H. performed bioinformatic analyses; V.R. obtained the magnetic resonance imaging images; F.L. evaluated the pathological findings; H.J.F. provided essential support and edited the paper; and H.-R.R. conceived and supervised the study, and wrote the paper.

Author Information Microarray data analysed and reported here have been deposited in the NCBI Gene Expression Omnibus⁴⁰ and are accessible through GEO accession number GSE56419. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-R.R. (hr.rodewald@dkfz.de).

METHODS

Mice. Recipient mice were C57BL/6, $Rag2^{-/-}\gamma_c^{-/-}Kit^{W/W^v}$, Kit^{W/W^v} , $Rag2^{-/-}\gamma_c^{-/-}$, $\gamma_c^{-/-}$, $Il7\alpha^{-/-}$, $Rag1^{-/-}$, $Rag1^{-/-}tg(OT-I)$ (all CD45.2⁺) as described^{17,41–44}. Histocompatible thymus donors were B6.SJL-*Ptpr*^{ca} Pep3^b/BoyJ (H-2^b; CD45.1⁺) (termed B6/Ly5.1), or F1(WB-Kit^{+/+} × B6/Ly5.1), or F1(WB-Kit^{+/+} × B6/Ly5.1-RAG2p-GFP transgenic¹⁸), or $Rag1^{-/-}$, or $Rag1^{-/-}tg(OT-I)$ mice. Both male and female mice were used at ages ranging from 6 to 20 weeks. No randomization and no blinding was used. All mouse experiments were approved by the Regierungspräsidien Tübingen or Karlsruhe.

Thymus transplantation and recipient analyses. Two thymus lobes were grafted as described⁴⁵ at the poles of the kidney of one recipient. Group sizes of approximately ten mice were used to detect leukaemia incidence rates with sufficient reliability. Assuming an incidence rate of 50% and allowing for a confidence interval width of 50%, that is, 50% ± 25%, 10 mice were required to control the type I error rate at 10%. With group sizes of at least 10 mice, we were able to detect a 4.5-fold increased risk as well as a 90% risk reduction for time to T-ALL development with 80% power in a two-group comparison based on the log-rank test, assuming a median time of 30 weeks in the standard group and a type I error rate of 5%. In the case of $Rag1^{-/-}$ recipient mice, only $n = 5$ were used (Fig. 2) in view of the fact that these hosts suppress thymus autonomy which is a prerequisite for T-ALL development. Although the $n = 8$ for $\gamma_c^{-/-}$ recipients is slightly below 10, the combined evidence obtained together with data from $Il7\alpha^{-/-}$ recipients ($n = 12$) underscores the role of the $Il7r$ in cell competition. Group sizes for control mice were greater than $n = 30$ to exclude even rare incidence of leukaemia in these strains. Survival curves were calculated with Prism 5.

Flow cytometry. Organs were harvested and single-cell suspensions prepared in PBS/5%FCS. Cells were blocked with 100 μ g ml⁻¹ mouse IgG (Jackson ImmunoResearch) for 15 min and stained for 30 min in an appropriately diluted antibody staining solution in PBS/5%FCS. Reagents used were CD3 allophycocyanin (APC)-Cy7 (17A2), CD4 phycoerythrin (PE)-Cy7 (GK1.5), CD8 APC (53-6.7), CD44 PerCP-Cy5.5 (IM7), CD45.1 PE (A20), CD45.1 PE-Cy7 (A20) (eBioscience), CD3 PE (145-2C11), CD4 PE, CD8 PE (53-6.7), CD11c PE (HL3), CD19 PE (1D3), CD25 PE (PC61), CD45.1 bio (A20), CD45.1 FITC (A20), CD45.2 PerCP-Cy5.5 (104), CD117 APC (2B8), NK1.1 PE (PK136), Gr1 PE (RB6-8C5), Ter119 PE (Ter119), Ki67 Alexa Fluor 647 (B56), Bcl2 PE (clone 100) (BD Pharmingen), CD11b PE (M1/70.15), Sca1 PE-Cy5.5 (Ly-6A/E), CD4 QDot605 (RM4-5) and Streptavidin QDot605 (Invitrogen/Molecular Probes). For definition of HSC, the lineage cocktail (Lin) was composed of CD3, CD4, CD8, CD11b, CD11c, Gr1, CD19, NK1.1, and Ter119. In the thymus, sorted populations were defined as follows: ETP (Lin⁻CD44⁺CD25⁻Kit⁺), TN2 (lin⁻CD44⁺CD25⁺), TN3a (CD4⁻CD8⁻CD3⁻CD44⁻CD25⁺CD28⁻), TN3b (CD4⁻CD8⁻CD3⁻CD44⁻CD25⁺CD28⁺), TN4 (CD4⁻CD8⁻CD3⁻CD44⁻CD25⁻CD28⁺), ISP (CD4⁻CD8⁺CD24⁺TCR β ⁻), DP (CD4⁺CD8⁺). For staining of blood cells, blood was collected from the tail vein directly into potassium-EDTA containing tubes (Sarstedt), and 500 μ l of PBS/5%FCS were added to dilute the blood. FICOLL-Paque PLUS (GE Healthcare) was added below the cell suspension. Samples were centrifuged at 250g for 20 min at room temperature, the interface was collected, and cells were washed with PBS/5%FCS. Staining was performed as indicated above. Annexin staining (Annexin V Apoptosis Detection Kit, eBioscience) was performed according to the manufacturers indications, with the antibodies used for the extracellular staining having been combined with annexinV in a 1-step staining.

Magnetic resonance imaging (MRI). 10 *Foxn1*^{nu/nu} (nude) mice with T-ALL and 2 healthy control nude mice were killed and immediately analysed in a dedicated high-field smallanimal MRI system (BioSpec 117/16, Bruker Biospin, Germany, Ettlingen) with a 4 cm quadrature volume resonator in coronal slice orientation. The MRI imaging protocol comprised a multi-slice T1-weighted RARE sequence with acquisition parameters as: TE/TR = 16 ms/600 ms, RARE factor = 8, $\Delta x/\Delta y/\Delta z = 150/150/1,000 \mu$ m³ with an additional inter-slice gap of 400 μ m. For volume quantification the data were transferred to a research software (Segment, MEDVISO AB, Sweden, Lund) and the spleen was manually contoured.

Histology and immunohistochemistry. Cryosections (5 μ m) of spleen and liver of T-ALL⁺ recipients, and healthy controls were air-dried, and either immediately used for haematoxylin and eosin (H&E) staining or dehydrated in acetone for 8 min, air dried, and stored at -80 °C for subsequent immunohistochemistry. H&E sections were stained for 8 min in haematoxylin (Merck), followed by rinse and wash for 5 min in tap water, and further stained in 0.25% eosin (Sigma) for 3 min. Slides were dipped into 75% ethanol to remove excess of eosin, kept for 1 min in 96% ethanol, 3 min in 99% ethanol (twice), 5 min in xylol (twice), and mounted with Entellan (Merck). For the immunohistochemistry of the liver, slides were rehydrated for 5 min in PBS and blocked in PBS 0.2% Tween 1% BSA 10% normal goat serum (Chemicon-Millipore) for 1 h at room temperature. Staining was performed in blocking buffer with diluted purified anti-CD8 antibody (53-6.7, Biolegend) overnight at 4 °C, followed by 3 washes in PBS 0.2% Tween,

and detection with goat anti-rat Cy3 (Biolegend) for 1 h at room temperature. Stained sections were washed, dried and mounted with Fluoromount G (Southern Biotech) containing 1.5 μ g ml⁻¹ Dapi (Roth). For immunohistochemistry of spleen sections, slides were rehydrated and blocked at room temperature for 30 min in PBS 2% BSA 0.1% NaN₃ 100 μ g ml⁻¹ mouse IgG (Jackson ImmunoResearch). Staining was performed in PBS 2% BSA 0.1% NaN₃ with diluted biotinylated-CD4 (GK1.5), CD8 (53-6.7), B220 FITC (RA3-6B2) and Ter119 APC (Ter-119) for 1 h at room temperature in a wet chamber, followed by rinse and wash in PBS for 10 min. Streptavidin Alexa594 staining followed for 1 h at room temperature in a wet chamber. Slides were washed in PBS for 10 min, and stained in 1 μ mol 4',6-diamidino-2-phenylindole (DAPI) in PBS for 10 min at room temperature. Stained sections were washed, and mounted with Fluoromount G (Southern Biotech).

Genomic DNA and array-CGH (aCGH). For preparation of genomic DNA, 5 × 10⁶ splenocytes were spun down, digested with proteinase K, and genomic DNA was purified with phenol-chloroform, and diluted in water. The DNA integrity was assessed by agarose gel electrophoresis and 2 μ g DNA were used for digestion with *AluI* and *RsaI*. The digested DNA was labelled with Cy5 (tumour DNA) or Cy3 (control DNA) with Genomic Enzymatic Labelling kit (Agilent Technologies), and quantified with a ND-1000 photometer (NanoDrop Technologies). Labelled DNA was hybridized to Agilent Mouse Genome aCGH Microarrays (Agilent Technologies) according to the manufacturer's guidelines. All steps following DNA purification were performed at Miltenyi Biotec GmbH.

For aCGH data analyses, raw intensity values were preprocessed using 'normexp' background correction and loess normalization, with consequent log₂ transformation of the data. Replicated probes were averaged, and probes with more than 10% missing values were removed. Missing values were imputed using a smoother to account for genomic structure. Circular binary segmentation (CBS) was used to detect change points in copy number states and to calculate smoothed intensities within segments. Copy number states that had close-by mean intensities were further merged after segmentation. Thresholds for defining copy number changes were set at ± 0.6 for log₂ smoothed intensities. All analyses were carried out using R 2.15/Bioconductor 2.11 with add-on packages limma, aCGH, snapCGH, DNACopy and cghMCR.

RNA, complementary DNA synthesis and RT-PCR. RNA was isolated using RNeasy B (IsoTex Diagnostics) according to the manufacturer's recommendations and dissolved in RNase free water. For cDNA synthesis, a DNase I (Invitrogen) digest step was performed, cDNA was synthesized with the transcript high fidelity kit (Roche) using oligo dT primers. Expression of canonical and alternative *Notch1* transcripts, as well as *Hes1* and β *actin*, was analysed by RT-PCR. For PCR amplification from cDNA, we used the following primers and conditions: *Notch1E1aE6* F: 5'-CCTCTCCCTCCAGGTCCATC-3', R: 5'-CAGGCATTGGGCATGAGCTGA-3', PCR product: ~1,300 bp, annealing temperature 60 °C; *Notch1E1E6* F: 5'-GCGAGCGCATGAAGGAACGA-3', R: 5'-CAGGCATTGGGCATGAGCTGA-3', PCR product: 1,115 bp, annealing temperature 65 °C; β *actin* F: 5'-CATC ACTATTGGCAACGAGC-3', R: 5'-ACGAGCTCAGTAACAGTCC-3', PCR product: 410 bp, annealing temperature 60 °C; *Hes1* F: 5'-ATGCCAGCTGATATA ATGGAG-3', R: 5'-ACGCTCGGGTCTGTGCTGAGC-3', PCR product: 317 bp, annealing temperature 60 °C.

Analyses of deletions and mutations in *Notch1*, and of TCR- β rearrangements. Genomic DNA was purified as described above. Genomic regions were amplified by PCR using the Expand Long Template PCR system (Roche). The amplified bands were cut from agarose gels, purified with GeneClean Spin Kit (MP Biomedicals), and either used for direct sequencing (deletion of the *Notch1* promoter), or cloned into pCR2.1-TOPO vector (Invitrogen) (*Notch1* exons 26, 27 and 34, and TCR- β rearrangements). At least 10 individual minipreps were sequenced for each cloned PCR product. For TCR- β , all analysed T-ALL samples had either one or two rearrangements. Sample #89 yielded three PCR products (Extended Data Fig. 3a). However, the two bigger PCR products corresponded to the same TCR sequence which was amplified by two different V β primers (one of which, V β 21, was cross reactive). Hence, also T-ALL #89 had only two distinct TCR- β rearrangements.

Deletions in the *Notch1* locus were analysed according to Ashworth *et al.*²⁹ by PCR on genomic DNA from T-ALL and from control DNA provided by T. Ashworth and J. Aster. Primers and PCR conditions used were: Δ Notch1 F: 5'-ATGGTGAATGCCTACTTTGTA-3', R: 5'-CGTTTGGGTAGAAGAGATGCTTTAC-3', PCR product: ~500 bp, annealing temperature 58 °C; *gp130* F: 5'-GGCTTTTCTCTGTTCTTG-3', R: 5'-CAGGAACATTAGGCCAGATG-3', PCR product: 615bp, annealing temperature 60 °C. Mutations in exons 26, 27 and 34 of the *Notch1* locus were analysed by PCR on genomic DNA as described by O'Neil *et al.*²⁴ using the following primers and PCR conditions: *Notch1* exon26 F: 5'-ACGGGAGGACCTAACCAAC-3', R: 5'-CAGCTTGGTCTCCAACACCT-3', PCR product: 792 bp, annealing temperature 60 °C; *Notch1* exon 27 F: 5'-CGCTGATGCTAAACACTGG-3', R: 5'-GTTTTGCCTGCATGTACGTC-3', PCR product: 505 bp, annealing temperature 60 °C; *Notch1* exon 34.1 F: 5'-GCTCCCTCATGTA

CCTCCTG-3', R: 5'-TAGTGGCCCCATCATGCTAT-3', PCR product: 904 bp, annealing temperature 62 °C; *Notch1* exon 34.2F: 5'-ATAGCATGATGGGGCC ACTA-3', R: 5'-CTTACCCTGACCAGGAAAA-3', PCR product: 893 bp, annealing temperature 62 °C. TCR rearrangements were analysed by PCR on genomic DNA using a pool of all *Vβ* (nomenclature according to <http://www.imgt.org>) forward primers with either *Jβ1.7* or *Jβ2.7* reverse primers. We used the following primers and PCR conditions:

Vβ primers: *Vβ1*: 5'-GCTGGAGCAAAACCAAGGTG-3', *Vβ2*: 5'-CTGGT GGCAGTCACAGGGA-3', *Vβ3*: 5'-GGGCAGATGGTGACCCTCAA-3'.

Vβ4: 5'-CCAAGGCGCTTCTACCTCA-3', *Vβ5*: 5'-GCCGCCAGAGCTC ATGTTTC-3', *Vβ12.1*, *Vβ12.2*, *Vβ12.3*: 5'-CTGGGGTTGTCCAGTCTCCA-3', *Vβ13.1*, *Vβ13.2*: 5'-GCTGCAGTCACCAAGCC-3', *Vβ13.3*: 5'-GCTGCAGT CACCCAAAGTC-3', *Vβ14*: 5'-GCTGGAGTCACCCAGTCTC-3', *Vβ15*: 5'-GC TGGAGTCACCCAGTCTC-3', *Vβ16*: 5'-GGCACAAGGTGACAGGGA-3', *Vβ17*: 5'-GCCTCTGCAGGCCTAGAGT-3', *Vβ19*: 5'-GGTACCGACAGGATTCAGGG-3', *Vβ20*: 5'-CCCATCAGTCATCCCAACTTAT-3', *Vβ21*: 5'-GGGGTTGTCCAG AATCCTAGACA-3', *Vβ23*: 5'-CCCTCCAGTCACTGTGCA-3', *Vβ24*: 5'-GG GCACACTGCCTTTTACTGTT-3', *Vβ26*: 5'-CTGAGGCAGGAGACTCAGCA-3', *Vβ29*: 5'-GGAATGTGGACAGGACATGAGC-3', *Vβ30*: 5'-CCAGTGTGTG GCTGACAGTCA-3'; *Jβ* primers: *Jβ1.7*: 5'-CCAAGACCATGGTCATCCAAC-3', *Jβ2.7*: 5'-TGAGAGCTGTCTCTACTATCGATT-3'. All *Vβ* (forward) primers were pooled for the PCR reactions and each was used at a final concentration of 0.05 μM. Reverse primers for either *Jβ1.7* or *Jβ2.7* were added to the reaction at a final concentration of 0.5 μM. PCR cycling conditions were: 94 °C for 2 min, 35 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 2 min followed by a final extension of 5 min at 72 °C.

Transcriptome analysis. For microarray experiments, at least 2 biological replicates per group were used. This limited group size was considered adequate to globally characterize groups and to detect strongest regulated features as groups were rather homogeneous. Because heterogeneity was unknown for the test groups, larger numbers of samples were analysed. RNA was isolated using RNeasy B (IsoTex Diagnostics). All samples were subject to quality control. cRNA was synthesized and hybridized to Illumina MouseWG-6 v2.0 BeadChips (at the Microarray Facility of the German Cancer Research Center).

Data were normalized and analysed with Chipster v2.9. A principal component analysis was applied so that principal components 1 and 2 explain 80% of the variance of the data.

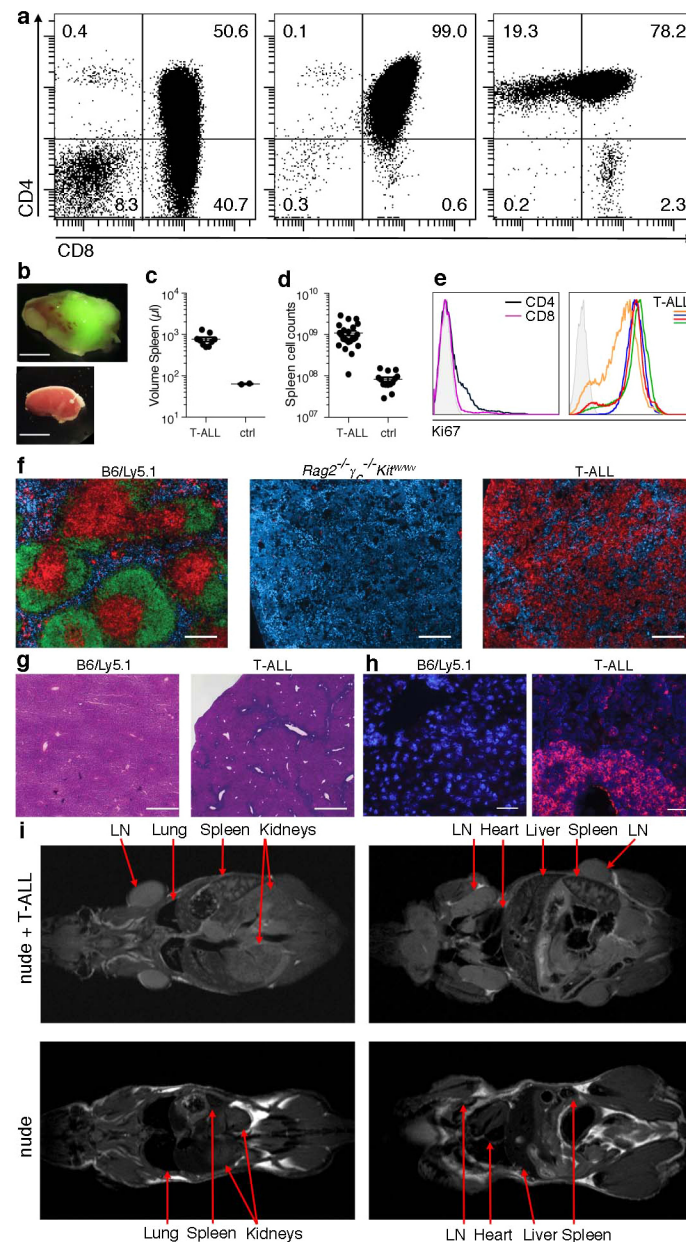
The clustering of the samples within groups when displaying principal component 1 versus 2, confirmed that the data could be grouped together for further analysis. Samples that did not cluster with their experimental groups were kept apart and not taken for subsequent analyses. This is the case of one thymus graft (shown in Fig. 4c) that presented higher similarities to normal thymus than to the

remaining thymus grafts. Another case is one of the 'old without competition' samples, that does not cluster with any of the remaining samples. Pairwise comparisons were carried out between transcriptomes of experimentally defined groups. Resulting lists of differentially expressed genes were overlaid in Venn diagrams and changes that were either specific or common between comparisons were identified. Ingenuity Pathway Analysis software was used to identify cancer-associated genes.

Comparison of the murine T-ALL signature to the transcriptomes of human T-ALL and human B-ALL.

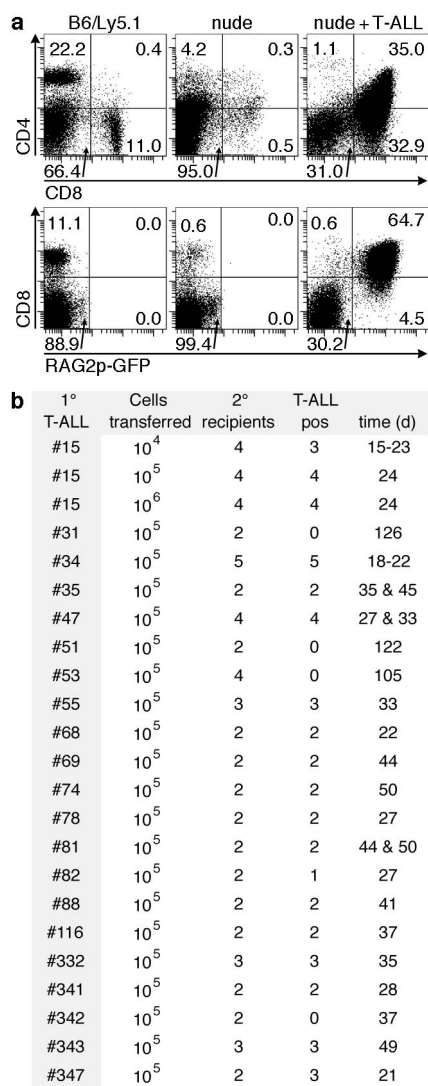
All statistical analyses were carried out using R, version 3.0.0. Published human transcriptome data sets were retrieved from the Gene Expression Omnibus data repository (<http://www.ncbi.nlm.nih.gov/geo/>) for T-ALL (GSE26713) and B-ALL (GSE13351). The complete data was normalized and pre-processed as described in the original work (ref. 31), and sample clustering was analysed using the method of average-linkage hierarchical clustering with Euclidean distance. The human orthologues of the murine genes displayed in Fig. 4f were retrieved from Ensembl (Ensembl 74, NCBI37), and most murine genes (59 out of 64) could be matched to the corresponding human genes. The mouse genes were mapped to 104 probe IDs in the mouse microarray data using R annotation package (illuminaMousev2.db, version 1.20.0), and the corresponding human orthologues were then mapped to 127 probe IDs using R annotation package (hgu133plus2.db, version 2.10.1). To perform correlation analysis, we computed a centeroid for each subtype of disease at two steps. First, probes were combined to genes based on Ensembl gene IDs, and each gene was represented by the intensity value of the probe with the highest variance across all samples in the data set (WGCNA R library, version 1.34). Second, the first principle component of the gene expression for each gene in each subtype was calculated. The Spearman rank correlation test was performed using the R library *stats* (version 3.0.0) and permutation *P* values were calculated using Spearman test of independence based on Monte-Carlo resampling (coin R library, version 1.0-23).

1. Cao, X. *et al.* Defective lymphoid development in mice lacking expression of the common cytokine receptor gamma chain. *Immunity* **2**, 223–238 (1995).
2. Peschon, J. J. *et al.* Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *J. Exp. Med.* **180**, 1955–1960 (1994).
3. Shinkai, Y. *et al.* RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
4. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
5. Rodewald, H. R., Kretzschmar, K., Swat, W. & Takeda, S. Intrathymically expressed c-kit ligand (stem cell factor) is a major factor driving expansion of very immature thymocytes *in vivo*. *Immunity* **3**, 313–319 (1995).



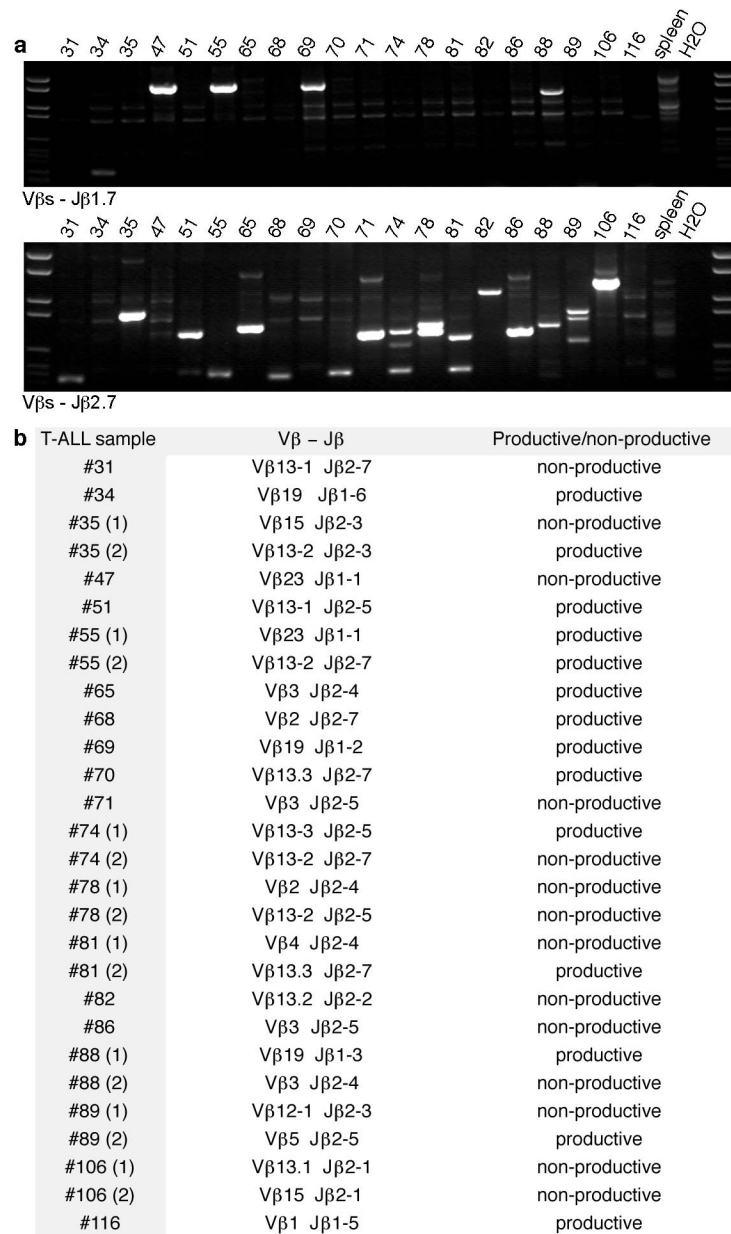
Extended Data Figure 1 | Characterization of T-ALL. **a**, Splenocytes from $Rag2^{-/-}\gamma_c^{-/-}Kit^{W/W^v}$ recipient mice with T-ALL were gated for CD45.1⁺ thymus donor cells, and analysed for CD4 and CD8 expression. All T-ALL samples were analysed in this manner. **b**, RAG2p-GFP reporter thymus-grafted (top), and non-transplanted (bottom) kidneys of a T-ALL⁺ recipient. Bright field and fluorescent images superimposed in the top panel (scale bars, 5 mm). All T-ALL⁺ mice had thymic tumours, and all RAG2p-GFP reporter thymus-derived T-ALL expressed Rag2 based on *in situ* fluorescence (shown here), or based on flow cytometry (Extended Data Fig. 5). **c**, Volume of spleens by magnetic resonance imaging (MRI). **d**, Spleen cellularity. **e**, Ki67 expression in CD4⁺ and CD8⁺ control splenocytes (left panel), and in T-ALL samples; each colour is an individual tumour (right panel). Control splenocytes or T-ALL cells unstained for Ki67 are shown in the grey shaded curves. **f**, Immunostaining of spleen cryosections of the indicated mice ($n = 2$ for T-ALL, and $n = 1$ for the controls) stained for T cells (CD4 plus CD8 in red), B cells (B220 in green), and erythrocytes (Ter119 in blue). Scale bar, 200 μ m. **g**, Liver cryosections of the indicated samples ($n = 4$ for T-ALL, and $n = 1$ for

the control) were stained with haematoxylin and eosin. Mice with T-ALL had perivascular and periportal infiltrates of a monomorphous small neoplasm that infiltrated into the adjacent liver parenchyma. At higher magnification (not shown), the neoplastic cells appeared medium-sized to large and showed a pronounced nuclear pleomorphism indicated by irregular edged nuclear contours with coarsely plumped chromatin. Cell borders were distinct, and a cytoplasmic rim was nearly invisible. These histomorphological findings were consistent with liver infiltration of a malignant haematopoietic neoplasm. Scale bar, 400 μ m. **h**, Liver cryosections of the indicated samples ($n = 2$ for T-ALL, and $n = 1$ for the control) were immunostained for CD8 (red) and counter-stained for DAPI (blue). Scale bars, 50 μ m. **i**, MRI analysis of T-ALL. Splenocytes from T-ALL⁺ mice were injected intravenously into athymic nude ($Foxn1^{nu/nu}$) mice. T-ALL⁺ nude mice were killed and immediately scanned by MRI ($n = 10$). Images shown are from a representative T-ALL⁺ mouse in two levels, revealing enlarged kidneys, liver, spleen and lymph nodes reflecting T-ALL dissemination (upper panels). Images from a nude mouse not receiving leukaemic cells ($n = 2$) are shown as controls (lower panels).



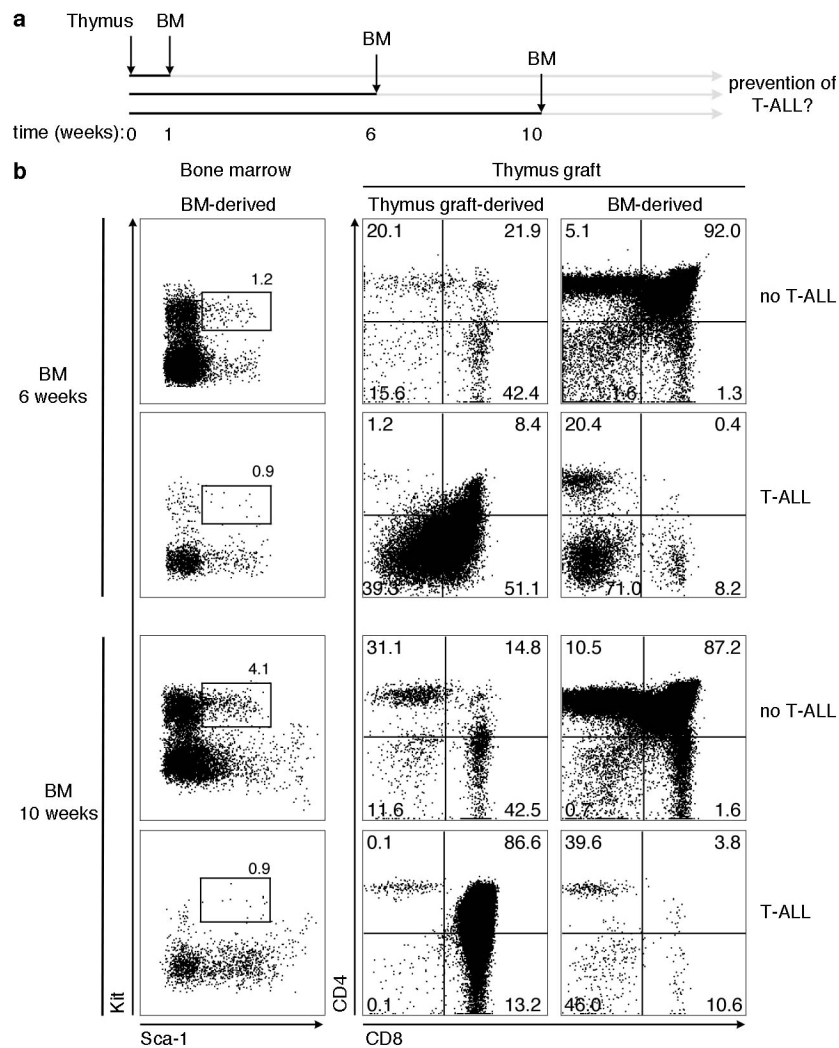
Average day of onset = 32 days

Extended Data Figure 2 | Transplantation of T-ALL into nude (*Foxn1^{nu/nu}*) mice. **a**, Analyses of CD4 and CD8 expression (top row), and CD8 and the RAG2p-GFP reporter (bottom row) in the spleens of a normal B6/Ly5.1 mouse (left), an unmanipulated nude mouse (middle), and a nude mouse 3 weeks after receiving 10⁵ splenocytes from a T-ALL⁺ *Rag2^{-/-}γc^{-/-}Kit^{W/W^v}* recipient (right). Data are representative for recipients listed in **b**. **b**, Summary of T-ALL sample numbers (primary (1°) T-ALL), numbers of cells transferred per nude mouse (cells transferred), numbers of nude mice injected per T-ALL sample (secondary (2°) recipients), numbers of nude mice developing T-ALL (T-ALL positive), and time in days until onset of the disease, or end of observation period (time).



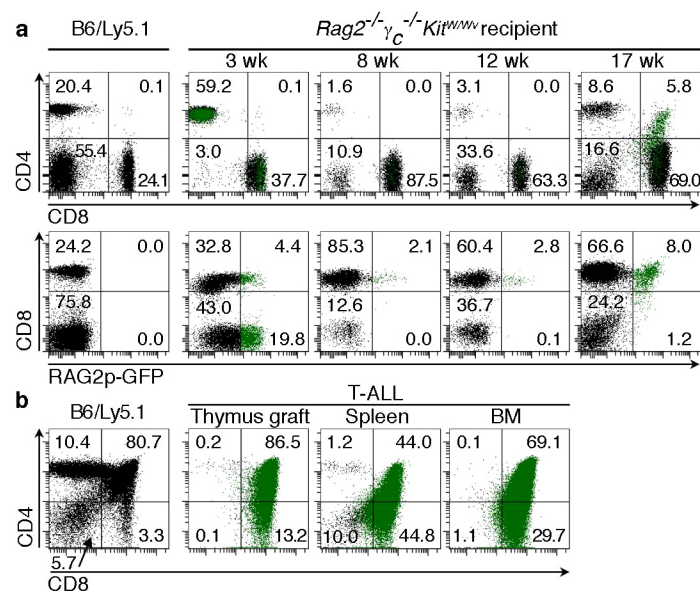
Extended Data Figure 3 | TCR-β rearrangements in T-ALL. **a**, PCR analyses of $V\beta$ to $J\beta1$ (top), and $V\beta$ to $J\beta2$ (bottom) rearrangements on genomic DNA from T-ALL⁺ spleen samples. TCR-β rearrangements were identified by PCR using a pool of $V\beta$ -gene-specific forward primers and reverse primers to either $J\beta1.7$ (top) or $J\beta2.7$ (bottom). Dominant, T-ALL-derived TCR-β

rearrangements were visualized by gel electrophoresis, and PCR products were cloned for DNA sequencing. **b**, T-ALL samples analysed (sample ID), identified $V\beta$ and $J\beta$ elements ($V\beta - J\beta$), and classification of in-frame, or out-of-frame rearrangements based on the reading frames of the obtained sequences (productive/non-productive).



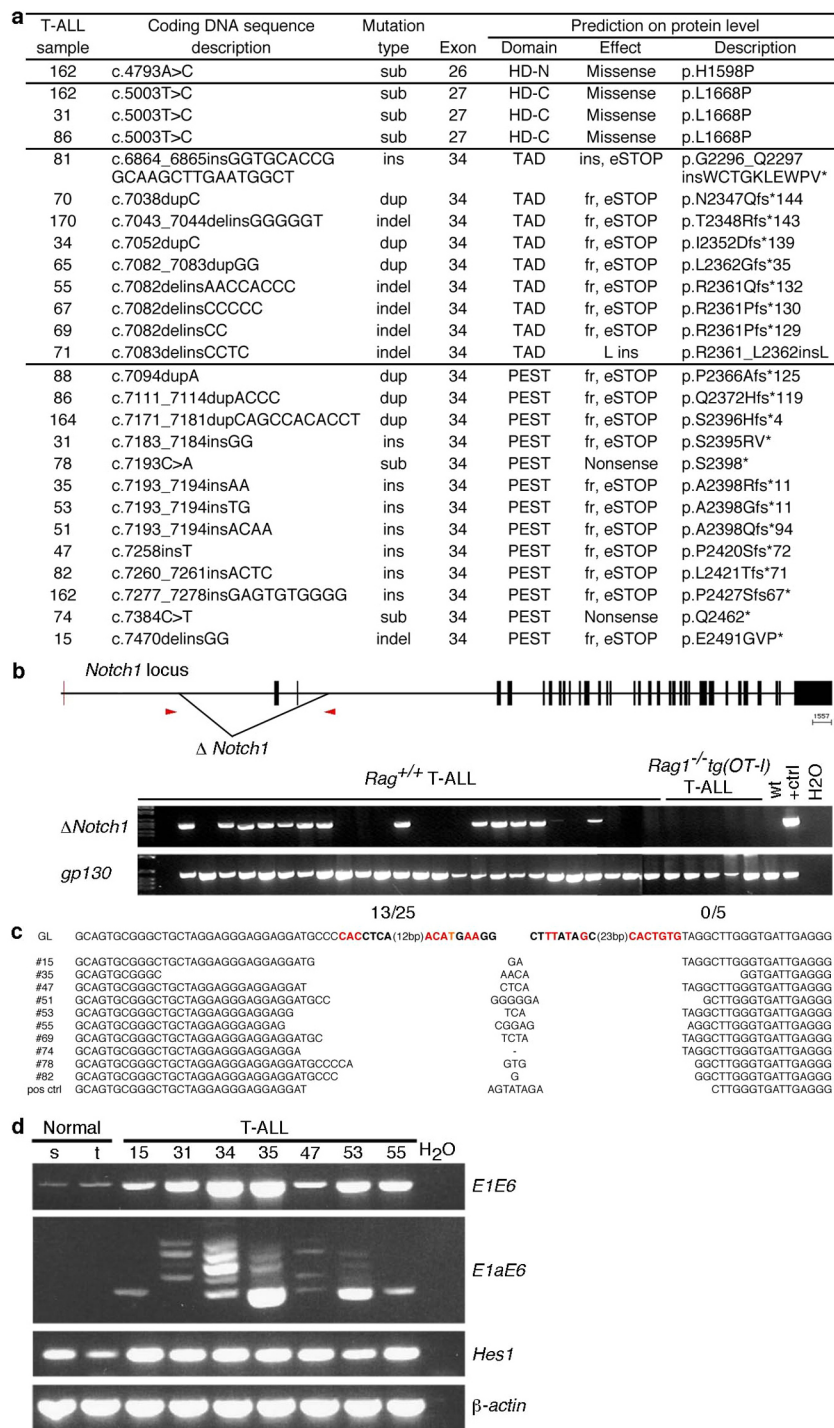
Extended Data Figure 4 | Experiments addressing prevention of T-ALL by reconstitution with functional wild-type bone marrow. **a**, Experimental design: *Rag2*^{-/-}*γc*^{-/-}*Kit*^{W/W^v} mice were transplanted with wild-type thymus (as described in Fig. 1a), and at 1, 6 or 10 weeks after thymus grafting were injected intravenously with 3×10^5 lineage-negative wild-type bone marrow cells (BM). T-ALL was prevented in mice receiving bone marrow 1 week after thymus transplantation (*n* shown in Fig. 2b). **b**, Flow cytometric analysis of bone marrow stem/progenitors demonstrating engraftment of donor 'rescue' bone marrow (left panels), and the cellular phenotypes in the thymus grafts (middle and right panels) (*n* shown in Fig. 2b). Bone marrow cells are pre-gated as lineage marker-negative (lin⁻) and donor cells (CD45.1⁺). Thymocytes are pre-gated for cells of graft origin (CD45.2⁺ CD45.1⁺, middle panels), and cells

of bone marrow 'rescue' origin (CD45.1⁺, right). Numbers in the plots refer to the percentage of gated cells. For each condition (bone marrow reconstitution at 6 or 10 weeks post thymus transplantation), examples are shown for ongoing T-cell development from the wild-type bone marrow, but not the thymus-resident cells, with no evidence for T-ALL (first and third rows), and for cases in which T-ALL developed while T-cell development of wild-type bone marrow origin did not prevail (second and fourth rows). The bone marrow data indicate that prevention of T-ALL was associated with stronger HSC engraftment by wild-type donor cells, however, it cannot be excluded that the overt T-ALL already suppressed haematopoiesis in the bone marrow, and hence the difference in the bone marrow may be secondary to the emergence of T-ALL.



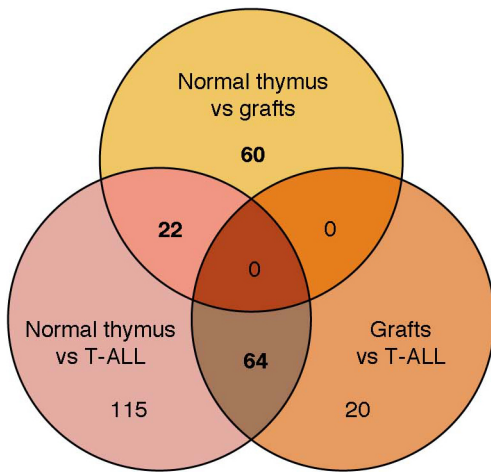
Extended Data Figure 5 | T-ALL emergence. **a**, Peripheral blood cells (gated for CD45.1⁺) from a normal control mouse (B6/Ly5.1), and a *Rag2*^{-/-} γ_c ^{-/-}*Kit*^{W/W^v} recipient carrying a RAG2p-GFP (CD45.1⁺) donor thymus were analysed by flow cytometry at the indicated time points.

GFP-positive cells are shown in green. **b**, The recipient shown in **a** was killed on week 18, and the indicated organs were analysed. RAG2p-GFP⁺-gated cells are green. Data are representative for RAG2p-GFP⁺ thymus-derived T-ALL ($n = 12$).

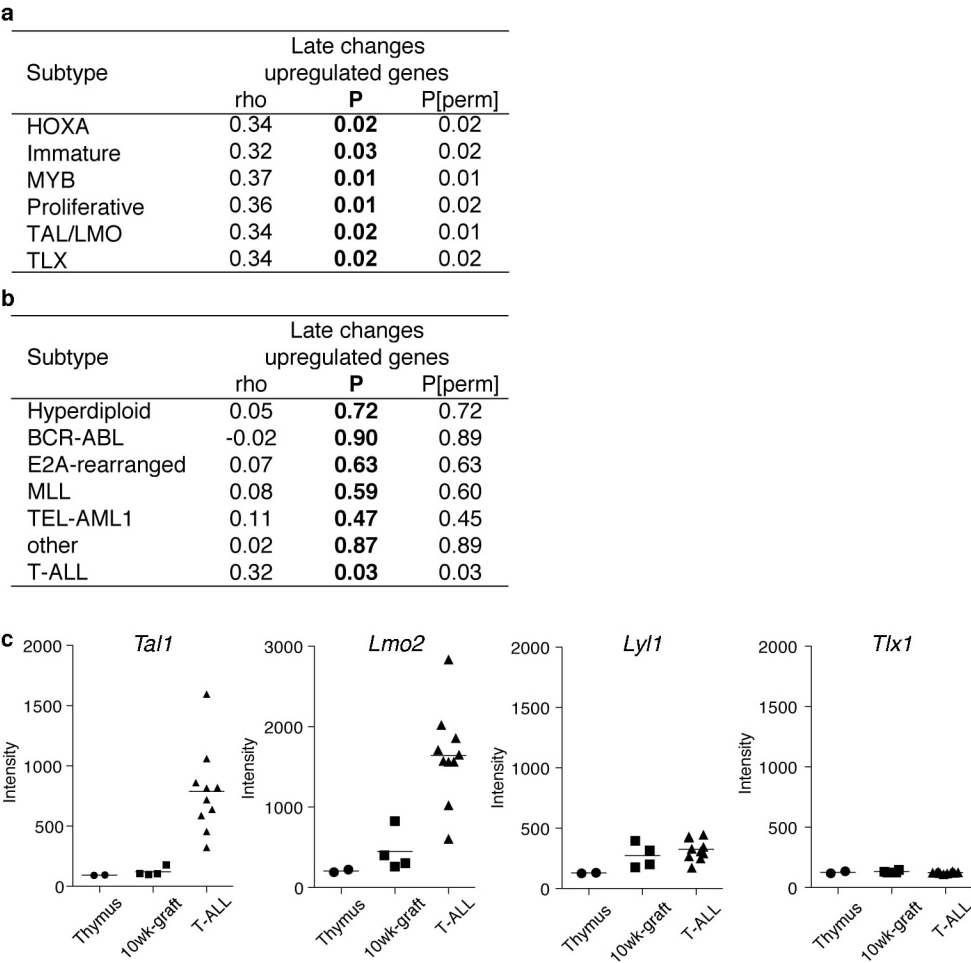


Extended Data Figure 6 | Mutations in and aberrant expression of *Notch1* in T-ALL. **a**, Exons 26, 27 and 34 of *Notch1* were sequenced and the mutation position, mutation type and predicted consequence for the protein are listed for all analysed T-ALL. All individual mutations were heterozygous and correspond to the triangles depicted in Fig. 3e. Substitutions (sub), insertions (ins), duplications (dup), complex deletions and insertions (indel), predicted frameshift (fr) and early STOP codons (eSTOP) are indicated, as well as the protein domains predicted to be affected: amino or carboxyl heterodimerization domains (HD-N or HD-C, respectively), transactivation domain (TAD), and PEST domain. **b**, Schematic view of the *Notch1* locus, depicting 5' deletions (Δ *Notch1*), and positions of the annealing sites for the primers used to detect deletions (red triangles). Exons are indicated by black bars. Genomic DNA from *Rag*^{+/+} (*n* = 25) and *Rag*^{-/-} (*n* = 5) T-ALL was analysed by PCR for the presence of the deletion²⁹. Amplification of the band with ~ 500 bp (Δ *Notch1*; upper gel) identifies the deletion. The non-deleted

locus is too large to yield a band; hence, absence of the band indicates no deletions. Glycoprotein 130 (*gp130*) was amplified as positive control for the presence of the DNA template. Deletions were seen in 13/25 *Rag*^{+/+}, but in 0/5 *Rag*^{-/-} T-ALL. **c**, PCR products were sequenced to characterize the junctions. The germline *Notch1* sequence is shown on the top (GL), revealing the sequence homology to the recombination signal sequence (RSS) highlighted in bold. The conserved residues in the putative cryptic RSS in the 5' region of the *Notch1* locus are shown in red. For 10 deletions, the sequences flanking the junctions are shown, with evidence of N nucleotide addition. The positive control DNA was derived from a T-ALL cell line that contained the deletion (provided by Ashworth and Aster). **d**, cDNAs from spleen (s) and thymus (t) from normal mice, and from T-ALL⁺ spleens were analysed for expression of canonical *Notch1* promoter transcripts (*E1E6*), alternative *Notch1* promoter transcripts (*E1aE6*), *Hes1*, and β -actin.

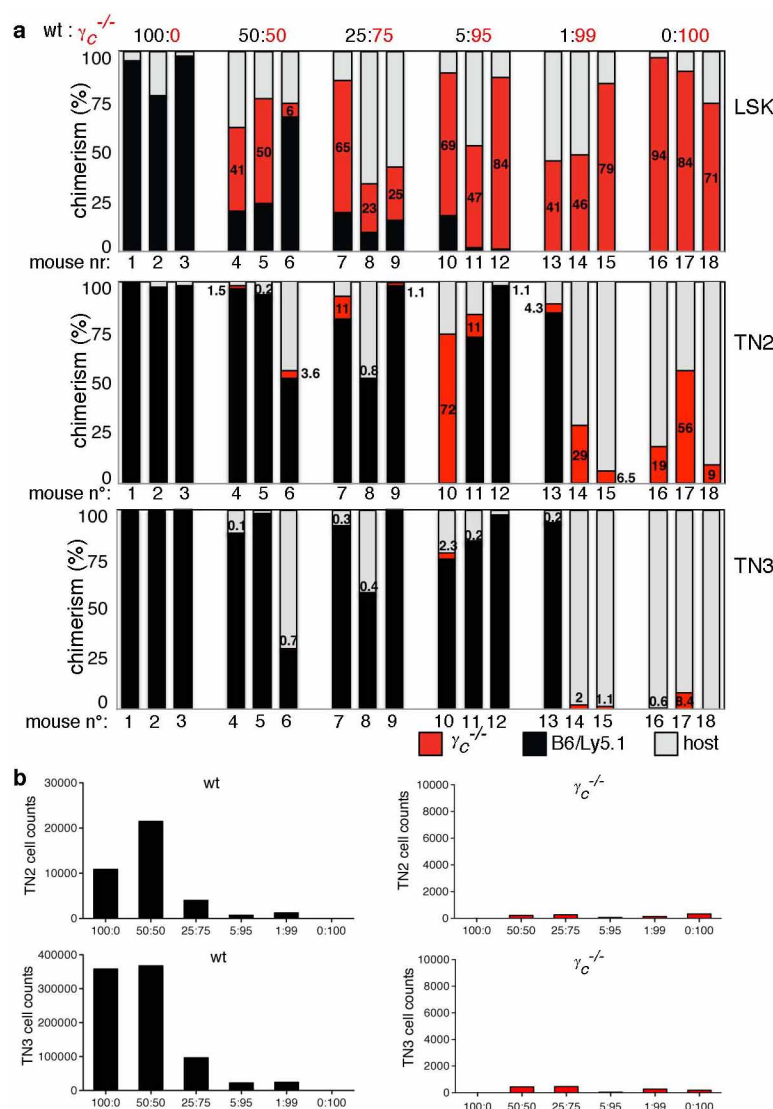


Extended Data Figure 7 | Venn diagrams. Diagrams underlay the transcriptome analysis in Fig. 4. The transcriptomes from normal thymus, thymus grafts (10 weeks after transplantation) and T-ALL samples were analysed by pairwise comparisons: normal thymus versus thymus grafts (top diagram); thymus grafts versus T-ALL (lower right diagram); normal thymus versus T-ALL (lower left diagram). The Venn diagram shows numbers of genes with expression changes of \geq fourfold for each comparison. The overlapping regions show the genes that were common to the different comparisons. The numbers 60, 22 and 64 (in bold) correspond to the genes displayed in the heat maps **d**, **e**, and **f**, respectively, in Fig. 4.



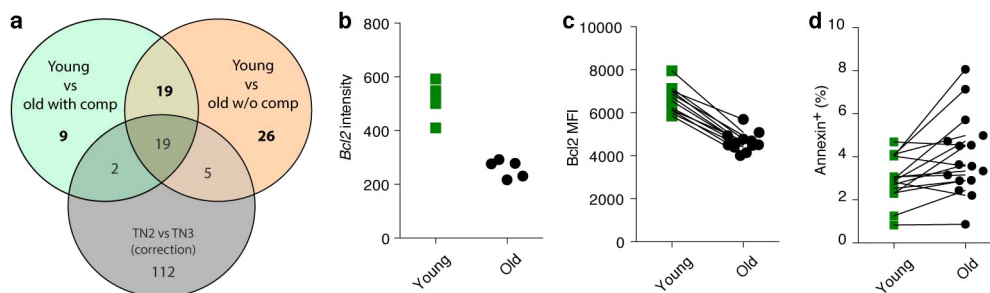
Extended Data Figure 8 | Molecular correlation between progenitor deprivation-driven murine T-ALL and human T-ALL. **a, b,** Genes upregulated in murine T-ALL (list from Fig. 4f) were compared to the transcriptome of human T-ALL (**a**), and human B-ALL (**b**). Upregulated changes are significantly and positively correlated with the expression of the orthologous genes all human T-ALL subtypes (GSE26713, 117 patient

samples) (**a**). Upregulated changes do not correlate with the expression of the orthologous genes in human B-ALL subtypes (GSE13351) (**b**). The 73 patient samples in GSE13351 also included 15 T-ALL samples that correlated with the mouse T-ALL samples (bottom row). **c,** Intensity values for the indicated genes were obtained from the normalized transcriptome data. Mean and standard deviation are displayed.



Extended Data Figure 9 | Sensitivity and stage of competition between wild-type and IL-7-unresponsive ($\gamma_c^{-/-}$) T-cell progenitors. a, A total of 3×10^5 lin^- donor bone marrow cells were either pure (100%) wild-type ($\gamma_c^{+/+}$) (CD45.1 $^+$; black bars) cells, or pure (100%) $\gamma_c^{-/-}$ (CD45.2 $^+$; red bars) cells, or mixtures of both genotypes (ratios indicated at the top), and used to reconstitute 1,100 rad-irradiated recipients (CD45.1 $^+$ CD45.2 $^+$). Each bar corresponds to a single mouse (no. 1–18). Percentages of chimaerism within bone marrow $\text{lin}^- \text{Sca-1}^+ \text{Kit}^+$ (LSK) (that contain haematopoietic stem and early progenitor cells) (top), within thymic TN2 (middle), and TN3 (bottom)

are displayed. Mice were analysed 4 weeks after bone marrow reconstitution. In the recipient bone marrow, the ratio of $\gamma_c^{+/+}$ versus $\gamma_c^{-/-}$ HSC donor engraftment closely mirrored the proportions of the input chimaerism. At the TN2 stage, few $\gamma_c^{-/-}$ T-cell progenitors were still detectable, but by the TN3 stage they were completely outcompeted by wild-type donor, or residual host progenitors. **b**, Cell numbers of TN2 (top) and TN3 (bottom) of wild type (left panels, black bars) or $\gamma_c^{-/-}$ (right panels, red bars) are displayed as averaged cell counts for each group of mice shown in **a**. Note that the scales of the y axis differ between the left and right panels.



Extended Data Figure 10 | Comparison of young versus old progenitors in the presence, or absence of competition: transcriptome, *Bcl2* mRNA and protein expression, and apoptosis. **a**, Venn diagram representing the overlap between pairwise comparisons of gene expression (threshold \geq twofold) between young and old with competition (green circle), and between young and old without competition (orange circle). To correct for potential developmental differences in gene expression at the TN2 to TN3 transition, we

sorted TN2 and TN3 cells from a normal thymus. The numbers 9, 26 (in Fig. 5 only 17/26 genes are displayed), and 19 (all in bold) correspond to the genes displayed in the heat maps in Fig. 5j, k and l, respectively. **b**, *Bcl2* mRNA levels as determined by fluorescence intensity in the expression array from sorted young and old TN2/TN3 cells. **c**, *Bcl2* protein levels, measured by intracellular FACS staining after gating on young and old TN3 cells, are depicted as mean fluorescence intensity (MFI). **d**, Percentage of annexin⁺ cells in TN3.

A Wolf–Rayet–like progenitor of SN 2013cu from spectral observations of a stellar wind

Avishay Gal–Yam¹, I. Arcavi¹, E. O. Ofek¹, S. Ben–Ami¹, S. B. Cenko², M. M. Kasliwal³, Y. Cao⁴, O. Yaron¹, D. Tal¹, J. M. Silverman⁵, A. Hoshesh⁴, A. De Cia¹, F. Taddia⁶, J. Sollerman⁶, D. Perley⁴, P. M. Vreeswijk¹, S. R. Kulkarni⁴, P. E. Nugent⁷, A. V. Filippenko⁸ & J. C. Wheeler⁵

The explosive fate of massive Wolf–Rayet stars¹ (WRs) is a key open question in stellar physics. An appealing option is that hydrogen-deficient WRs are the progenitors of some hydrogen-poor supernova explosions of types IIb, Ib and Ic (ref. 2). A blue object, having luminosity and colours consistent with those of some WRs, has recently been identified in pre-explosion images at the location of a supernova of type Ib (ref. 3), but has not yet been conclusively determined to have been the progenitor. Similar work has so far only resulted in non-detections⁴. Comparison of early photometric observations of type Ic supernovae with theoretical models suggests that the progenitor stars had radii of less than 10^{12} centimetres, as expected for some WRs⁵. The signature of WRs, their emission line spectra, cannot be probed by such studies. Here we report the detection of strong emission lines in a spectrum of type IIb supernova 2013cu (iPTF13ast) obtained approximately 15.5 hours after explosion (by ‘flash spectroscopy’, which captures the effects of the supernova explosion shock breakout flash on material surrounding the progenitor star). We identify Wolf–Rayet-like wind signatures, suggesting a progenitor of the WN(h) subclass (those WRs with winds dominated by helium and nitrogen, with traces of hydrogen). The extent of this dense wind may indicate increased mass loss from the progenitor shortly before its explosion, consistent with recent theoretical predictions⁶.

Wolf–Rayet stars are massive stars stripped of their outer, hydrogen-rich envelope. These stars blow strong, hydrogen-poor winds. The inner part of the wind engulfing the star is dense and optically thick, and efficiently absorbs the ionizing continuum from the hot stellar surface. Farther from the star, the density drops and the wind becomes optically thin in the continuum, leading to a rich emission spectrum of recombination lines. Detailed models of such spectra can be constructed^{1,7} and depend essentially on only three parameters: the effective temperature, T_{eff} , in the line-forming region, a normalized radius, R_t (a combination of the stellar radius, luminosity, mass loss rate and wind terminal velocity⁷), and the chemical composition, Z , of the wind (assumed to be uniform, spherical and of a constant mass loss rate). The composition of the wind determines Wolf–Rayet spectral classes: stars with dominant He and N lines belong to the WN class (with those also showing traces of H usually denoted as WN(h)), stars with strong carbon lines belong to the WC class and rare (and possibly hotter) stars with oxygen-rich spectra belong to the WO class¹.

Shortly after a WR explodes as a supernova, the outer parts of the wind (which in some cases¹ extend to radii of $>10^{13}$ cm) that have not yet been swept up by the expanding supernova ejecta will emit strong recombination lines in response to ionizing flux released by the explosion shock breakout from the stellar surface. We estimate an increase in ionizing luminosity by a factor of order 10^2 – 10^4 , assuming an initial absolute magnitude range of $-2.5 \text{ mag} < M < -10 \text{ mag}$ for the exploding WR¹

and a typical early-time luminosity of $M = -12.5 \text{ mag}$ for the resulting supernova^{3,5,8}. The effective temperature, T_{eff} , will also change, being very high ($>10^5 \text{ K}$) shortly after explosion⁹ and decreasing as the shocked supernova ejecta cool. The radiation illuminating the surviving Wolf–Rayet wind will thus effectively vary monotonically through the range of temperatures in WR line-forming regions. Because the wind parameters (composition, mass loss rate and terminal velocity) do not change, the measured line spectrum observed shortly after explosion should be similar to that of a WR with the spectral class of the exploding progenitor (because the spectral classes mainly reflect the wind composition). The high wind densities around WRs (with electron densities of $n_e = 10^{11}$ – 10^{12} cm^{-3}) imply short recombination times¹⁰, $t_{\text{rec}} \approx 3.9 \times 10^{12} (n_e / \text{cm}^{-3})^{-1} (T / 10^4 \text{ K})^{0.85} \text{ s}$, typically of a few minutes for Wolf–Rayet densities and temperatures, and so the emitted spectrum will promptly react to the rapidly evolving supernova radiation field.

We obtained rapid spectroscopic observations of the recent type IIb supernova SN 2013cu (iPTF13ast) shortly after shock breakout (by flash spectroscopy; see Methods). This event was first detected by the Intermediate Palomar Transient Factory (iPTF) survey¹¹ on 2013 May 3.18 (we express dates in coordinated universal time (UTC)), photometrically confirmed 5.8 h later and promptly identified by an on-duty astronomer who triggered rapid follow-up observations¹², including an optical spectrum obtained just 4 h later. Analysis of the early-time light curve of this supernova (Extended Data Fig. 1) suggests that it exploded on May 2.93, implying that the first iPTF detection and the first spectrum correspond to only 5.7 h and 15.5 h after the explosion, respectively. A full description of the supernova and its evolution will be reported in a forthcoming publication (A.G.-Y. *et al.*, manuscript in preparation). We note that this event was independently observed by the MASTER survey on May 5.3 ($\sim 2.3 \text{ d}$ after explosion), and it was assigned the name SN 2013cu following spectroscopic confirmation¹³.

Our first spectrum of SN 2013cu reveals a continuum and emission lines that bear a striking resemblance to spectra of WRs (Fig. 1a). According to accepted Wolf–Rayet terminology¹, the spectrum is classified as subclass WN6(h) (Fig. 1a, blue trace); the relative strength of nitrogen to carbon lines precludes a WC classification, and the absence of any high-excitation oxygen lines is inconsistent with a WO star. The stronger lines (H α , H β , N IV $\lambda 7,115$ (7,115 Å wavelength) and He II $\lambda 5,411$) exhibit a complex profile (Fig. 2) consisting of a relatively broad base ($\sim 2,500 \text{ km s}^{-1}$ full width at zero intensity (FWZI)) on which prominent narrow, unresolved lines (FWZI $\approx 3\text{Å}$; velocity dispersion, $<150 \text{ km s}^{-1}$) are superimposed. This is consistent with predictions for Wolf–Rayet pre-supernova wind velocities^{14,15}, although we cannot exclude the possibility that at least some of the observed line broadening is produced by electron scattering rather than genuine velocity dispersion. To the best of our knowledge, no similar spectra of a stripped (H-poor) supernova have been acquired previously. Wolf–Rayet-like

¹Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot 76100, Israel. ²Astrophysics Science Division, NASA Goddard Space Flight Center, Mail Code 661, Greenbelt, Maryland 20771, USA. ³Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, California 91101, USA. ⁴Cahill Center for Astrophysics, California Institute of Technology, Pasadena, California 91125, USA. ⁵Department of Astronomy, University of Texas, Austin, Texas 78712, USA. ⁶The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, 10691 Stockholm, Sweden. ⁷Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁸Department of Astronomy, University of California, Berkeley, California 94720-3411, USA.

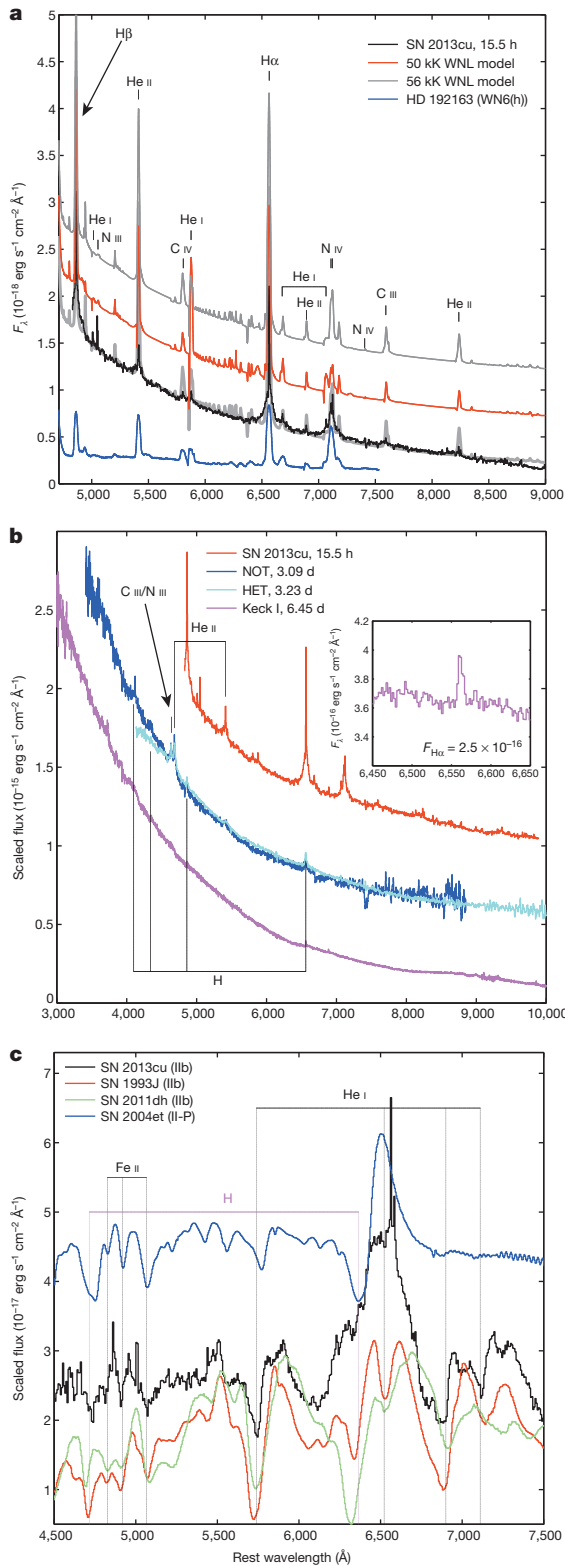


Figure 1 | Spectroscopy of the type IIb supernova SN 2013cu reveals transient Wolf–Rayet-like features. **a**, The early spectrum of SN 2013cu (black; specific flux, F_λ , versus wavelength) is compared with models of WNL-class WRs⁷ (red and grey), showing remarkable similarity in continuum shape and in line features (strong He, N and Balmer lines indicate a WN6(h) classification; compare with HD 192163 (blue)). **b**, Emission line evolution during the first week. The first spectrum (red) is compared with later spectra. By day 3 (blue and cyan), the initially strong Wolf–Rayet features disappear, but the $H\alpha$ line remains constant at $3.4 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. The spectrum on day 6 (magenta) is almost featureless, except for weak $H\alpha$ emission (inset) with an intensity less than a tenth that of day 3, probably because the line-forming region has been cleared by the expanding ejecta. **c**, SN 2013cu is of type IIb. A spectrum of SN 2013cu 69 days after explosion (black) is compared with the prototypical type IIb supernovae SN 1993J (60 days; red) and SN 2011dh (43 days; green), and with the typical non-stripped type II-P supernova SN 2004et (45 days; blue). SN 2013cu, SN 2011dh and SN 1993J exhibit strong blue-shifted He I absorption at 5,876, 6,678, 7,065 and 7,281 Å (marked with black vertical lines), which are not detected in type II-P supernovae. See Methods for more details.

around 50,000 K. This temperature is consistent with the lower limit obtained from early ultraviolet photometry from NASA’s Swift spacecraft (Extended Data Fig. 1). The essentially perfect match of the observed and modelled continuum shapes indicates that dust reddening must be negligible; any pre-existing dust must have been destroyed by the supernova explosion flash (Methods). We note that among the large catalogue of Galactic WN stars specifically modelled in this manner⁷, no stars drive winds that require more than 56% H by mass. Presumably, custom spectral fits⁷ could be calculated and used to determine more accurately the physical parameters of the detected Wolf–Rayet wind. Assuming that the spectrum was obtained 15.5 h after explosion and that a standard ejecta velocity of 10^4 km s^{-1} is applicable, the narrow-line-emitting material must be located at radii greater than $\sim 5 \times 10^{13} \text{ cm}$ for it not to have been swept up and accelerated by the expanding ejecta. This lower limit is consistent with the extent of some Wolf–Rayet winds, where the line-formation region extends out to several (five to ten) times¹⁷ the hydrostatic radius. Recent pre-supernova Wolf–Rayet models¹⁴ suggest hydrostatic radii of 10–20 solar radii for WN supernova progenitors, consistent with line-formation regions extending to several hundred solar radii¹, or $>10^{13} \text{ cm}$.

We further constrain the physical location of the wind using the following calculation. We measure the $H\alpha$ line flux from our spectrum (calibrated to our host-subtracted photometry) and obtain $F_{H\alpha} = 3.4 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. We translate this to line luminosity using a luminosity distance to the host galaxy, UGC 9379, of $d = 108 \text{ Mpc}$ (calculated for a flat Λ cold dark matter cosmology with Hubble constant $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$, a matter density of $\Omega_m = 0.27$ and a redshift $z = 0.025734$, obtained from the NASA Extragalactic Database (NED), as well as negligible extinction), obtaining $L_{H\alpha} = 4.8 \times 10^{39} \text{ erg s}^{-1}$. We can then estimate¹⁸ the pre-explosion H mass loss rate

$$\dot{M} = 0.01 (L_{H\alpha} / (2 \times 10^{39} \text{ erg s}^{-1}))^{1/2} \times (v_w / (500 \text{ km s}^{-1})) (r / (10^{15} \text{ cm}))^{1/2} M_\odot \text{ yr}^{-1}$$

assuming that the lines are formed at a radius r by recombination, that the width of the emitting shell is similar to its radius, that the explosion is spherical symmetric and that the density of the wind decreases as r^{-2} . We assume a wind velocity of $v_w = 2,500 \text{ km s}^{-1}$, consistent with our spectra and as expected for WRs, but our results are not sensitive to this value and remain essentially unchanged for velocities in the range $100 < v_w < 2,500 \text{ km s}^{-1}$.

We check for self-consistency by calculating the electron density and, hence, the Thomson optical depth

$$\tau = 0.3 (\dot{M} / (0.01 M_\odot \text{ yr}^{-1})) \times (v_w / (500 \text{ km s}^{-1}))^{-1} (r / (10^{15} \text{ cm}))^{-1}$$

spectroscopic features have been observed in spectra of some H-rich (non-stripped) supernovae obtained at substantially later epochs, and their typically much broader lines were interpreted as emerging from interaction with circumstellar material¹⁶ (CSM). We further discuss these previous observations in Extended Data Fig. 3.

We analyse our very early spectrum using the PoWR grid of Wolf–Rayet spectral models⁷ (<http://www.astro.physik.uni-potsdam.de/~wrh/PoWR/powrgrid1.html>). We find an excellent fit with WN(h) models calculated assuming an H fraction of 20% (by mass) and a temperature of

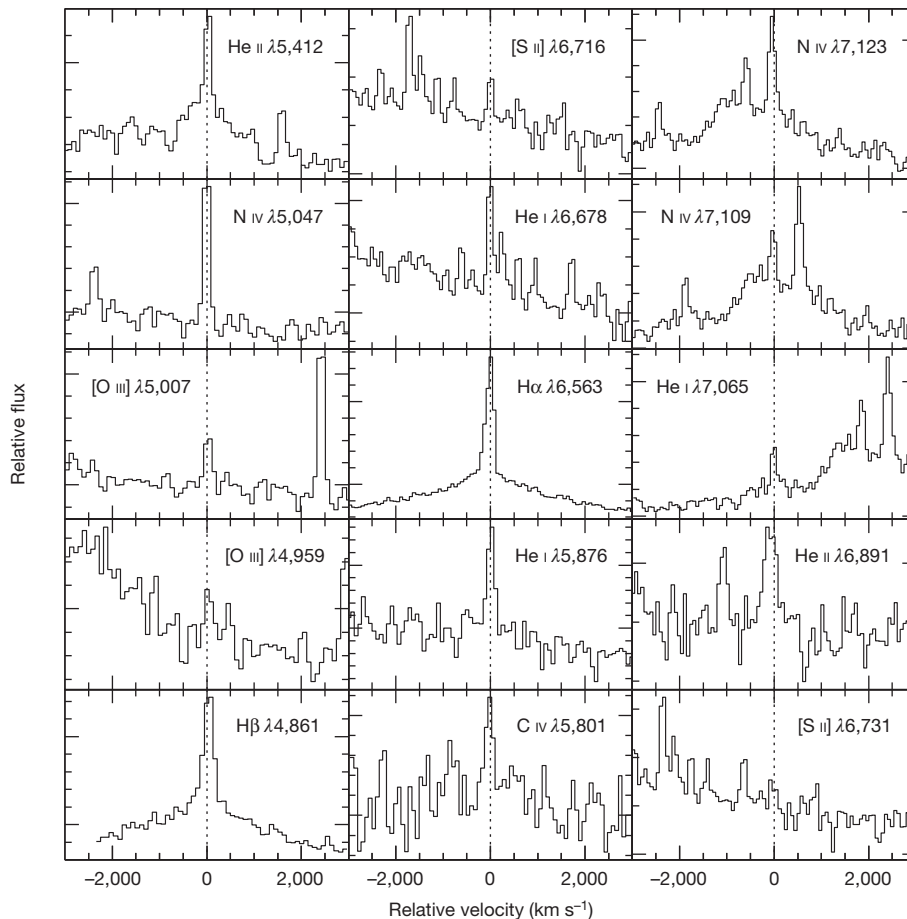


Figure 2 | Emission line velocity structure at 15.5 h. The strongest lines (H α , H β , He II λ 5,411 and the N IV λ 7,115 complex) show broad wings

extending out to $\sim 2,500 \text{ km s}^{-1}$. Other weaker lines are narrow and unresolved.

expected at this radius, and require it to be lower than $\tau = 1$ for the line emission to escape. We find that this self-consistency requirement places a lower limit of $r = 2 \times 10^{14} \text{ cm}$ on the radius of the line-formation region, and implies substantial mass loss rates, $\dot{M} > 0.03 M_{\odot} \text{ yr}^{-1}$. If we interpret the disappearance of essentially all emission lines from our day-6 spectrum (Fig. 1b) as evidence that the wind was swept up by the expanding ejecta (moving at 10^4 km s^{-1}), the radius of the line-emitting region must be $r < 5.2 \times 10^{14} \text{ cm}$, which is fully consistent with our estimates.

We can then calculate the total H mass by integrating over r :

$$M_{\text{tot}} = 0.006 \left(\dot{M} / (0.01 M_{\odot} \text{ yr}^{-1}) \right) \times (v_w / (500 \text{ km s}^{-1}))^{-1} (r / (10^{15} \text{ cm}))^{-1} M_{\odot}$$

This indicates a range of $0.008 M_{\odot} < M_{\text{tot}} < 0.0035 M_{\odot}$ for the range of permitted H masses. Assuming that the typical H abundance for WN(h) stars ($\sim 20\%$) applies, the total wind mass (dominated by He) can be estimated to be several times larger than these values. Detailed simulations¹⁹ show that as little as $0.1 M_{\odot}$ of He-dominated CSM would result in strong spectroscopic interaction signatures (that we do not observe), consistent with our derived total masses.

We conclude that we have directly detected a Wolf-Rayet-like wind from the supernova progenitor with a WN(h) spectral class, indicating a low H mass fraction. Assuming that the wind composition we measure represents the surface composition of the progenitor star, our observations indicate that some members of the spectroscopic WN(h) Wolf-Rayet class explode after having lost most of their hydrogen envelope, exposing the CNO-processed, N-rich He layer below. Analysis of photometric and spectroscopic follow-up observations (A.G.-Y. *et al.*, manuscript in preparation) confirms that the explosion was indeed a

supernova of type IIb (Fig. 1c), as expected if the progenitor was a massive star that lost all but $\sim 0.1 M_{\odot}$ of its H envelope²⁰.

Our observations have interesting implications. First, we note that the derived values of the mass loss rate and emission-line-region size are quite extreme compared with known Wolf-Rayet observations and radiatively driven models²¹, including models with clumpy, inflated atmospheres²². This suggests that the mass loss rate from the progenitor star may have increased shortly (of order 1 yr for the assumed velocities) before its explosion. Interestingly, such pre-supernova activity may be explained by recent wave-driven models⁶, or a more extreme envelope inflation²² may be indicated. These data can thus provide a key diagnostic of the final stages of nuclear core burning in massive stars, which are currently poorly understood, with possible implications for the explosion mechanism itself. In any case, the star probably exploded inside a thick wind, and the explosion shock may have broken out from the opaque inner wind rather than from the hydrostatic surface of the star⁹.

Our finding is in general accord with some previous work on type IIb supernova progenitors. Direct imaging of the progenitor of SN 2008ax (ref. 23) is consistent with a WN(h) progenitor. Furthermore, increased mass loss during the final year before explosion may inflate the apparent photospheric radius of the pre-supernova star, making stars with compact cores appear to have extended (low-mass) envelopes²² and possibly reconciling the conflicting findings about the progenitor of the type IIb supernova SN 2011dh (refs 24–27). Regardless of the exact mechanism, our observations suggest that substantial Wolf-Rayet-like winds pre-date at least some type IIb supernovae. A strong metallicity dependence of this process may explain the trend in the type IIb/type Ib supernova number ratio with host-galaxy metallicity²⁸. Future studies of numerous additional supernova progenitors via their spectroscopic

wind signatures (Methods) would provide powerful constraints on the final stages of massive-star evolution.

METHODS SUMMARY

Photometry. The iPTF survey telescope was used to obtain *r*-band observations. Photometry was measured using our custom pipeline performing point spread function (PSF) photometry on iPTF images after removing a reference image constructed from pre-explosion data using image subtraction. Swift ultraviolet absolute AB magnitudes (Extended Data Fig. 1) were measured using standard pipeline reduction and were corrected for host-galaxy contamination using late-time Swift images. **Spectroscopy.** Our earliest (15.5 h) and latest (69 d) spectra were obtained using the DEIMOS spectrograph mounted on the Keck II 10-m telescope using the grating with 600 lines per millimetre and an exposure time of 600 s. Additional spectra were obtained using ALFOSC mounted on the 2.56-m NOT telescope, LRS mounted on the 10.4-m HET telescope (day 3), and LRIS mounted on the Keck I 10-m telescope (day 6). All spectra were reduced using standard pipelines and are digitally available on WISEREP²⁹. The method of flash spectroscopy is described in detail in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 October 2013; accepted 25 March 2014.

- Crowther, P. A. Physical properties of Wolf-Rayet stars. *Annu. Rev. Astron. Astrophys.* **45**, 177–219 (2007).
- Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).
- Cao, Y. *et al.* Discovery, rise and progenitor of a stripped envelope supernova iPTF13bvn. *Astron. Astrophys. J.* **775**, L7 (2013).
- Eldridge, J. J., Fraser, M., Smartt, S. J., Maund, J. R. & Crockett, R. M. The death of massive stars - II. Observational constraints on the progenitors of type Ibc supernovae. *Mon. Not. R. Astron. Soc.* **436**, 774–795 (2013).
- Corsi, A. *et al.* Evidence for a compact Wolf-Rayet progenitor for the type Ic supernova PTF 10vgv. *Astron. Astrophys. J.* **747**, L5 (2012).
- Shiode, J. H. & Quataert, E. Setting the stage for circumstellar interaction in core-collapse supernovae II: wave-driven mass loss in supernova progenitors. *Astron. Astrophys. J.* **780**, 96 (2014).
- Hamann, W.-R., Gräfener, G. & Liermann, A. The Galactic WN stars. Spectral analyses with line-blanketed model atmospheres versus stellar evolution models with and without rotation. *Astron. Astrophys.* **457**, 1015–1031 (2006).
- Dessart, L. *et al.* Core-collapse explosions of Wolf-Rayet stars and the connection to type IIb/Ib/Ic supernovae. *Mon. Not. R. Astron. Soc.* **414**, 2985–3005 (2011).
- Ofek, E. O. *et al.* Supernova PTF 09uj: a possible shock breakout from a dense circumstellar wind. *Astron. Astrophys. J.* **724**, 1396–1401 (2010).
- Osterbrock, D. E. & Ferland, G. J. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* 47 (University Science Books, 2006).
- Law, N. M. *et al.* The Palomar Transient Factory: system overview, performance, and first results. *Publ. Astron. Soc. Pacif.* **121**, 1395–1408 (2009).
- Gal-Yam, A. *et al.* Real-time detection and rapid multiwavelength follow-up observations of a highly subluminal type II-P supernova from the Palomar Transient Factory survey. *Astron. Astrophys. J.* **736**, 159 (2011).
- Denisenko, D. *et al.* Supernova 2013cu in UGC 9379 = PSN J14335897+4014207. *Central Bur. Electr. Telegr. abstr.* 3540 (2013).
- Groh, J. H., Meynet, G., Georgy, C. & Ekström, S. Fundamental properties of core-collapse supernova and GRB progenitors: predicting the look of massive stars before death. *Astron. Astrophys.* **558**, A131 (2013).
- Niedzielski, A. & Skorzynski, W. Kinematical structure of Wolf-Rayet winds. I. Terminal wind velocity. *Acta Astron.* **52**, 81–104 (2002).
- Leonard, D. C., Filippenko, A. V., Barth, A. J. & Matheson, T. Evidence for asphericity in the type IIin supernova SN 1998S. *Astron. Astrophys. J.* **536**, 239–254 (2000).
- Hillier, D. J. An empirical model for the Wolf-Rayet star HD 50896. *Astron. Astrophys. J.* **63** (suppl.), 965–981 (1987).
- Ofek, E. O. *et al.* SN 2009ip: constraints on the progenitor mass-loss rate. *Astron. Astrophys. J.* **768**, 47 (2013).
- Chugai, N. N. Circumstellar interaction in type Ibc supernovae and SN 2006jc. *Mon. Not. R. Astron. Soc.* **400**, 866–874 (2009).
- Hachinger, S. *et al.* How much H and He is 'hidden' in SNe Ib/c? - I. Low-mass objects. *Mon. Not. R. Astron. Soc.* **422**, 70–88 (2012).
- Vink, J. S. *et al.* Wind modelling of very massive stars up to 300 solar masses. *Astron. Astrophys.* **531**, A132 (2011).
- Gräfener, G., Owocki, S. P. & Vink, J. S. Stellar envelope inflation near the Eddington limit. Implications for the radii of Wolf-Rayet stars and luminous blue variables. *Astron. Astrophys.* **538**, A40 (2012).
- Crockett, R. M. *et al.* The type IIb SN 2008ax: the nature of the progenitor. *Mon. Not. R. Astron. Soc.* **391**, L5–L9 (2008).
- Arcavi, I. *et al.* SN 2011dh: discovery of a type IIb supernova from a compact progenitor in the nearby galaxy M51. *Astron. Astrophys. J.* **742**, L18 (2011).
- Bersten, M. C. *et al.* The type IIb supernova 2011dh from a supergiant progenitor. *Astron. Astrophys. J.* **757**, 31 (2012).
- Van Dyk, S. D. *et al.* The progenitor of supernova 2011dh has vanished. *Astron. Astrophys. J.* **772**, L32 (2013).
- Ergon, M. *et al.* Optical and near-infrared observations of SN 2011dh - the first 100 days. *Astron. Astrophys.* **562**, A17 (2014).
- Arcavi, I. *et al.* Core-collapse supernovae from the Palomar Transient Factory: indications for a different population in dwarf galaxies. *Astron. Astrophys. J.* **721**, 777–784 (2010).
- Yaron, O. & Gal-Yam, A. WISEREP—an interactive supernova data repository. *Proc. Astron. Soc. Pacif.* **124**, 668–681 (2012).

Acknowledgements This research was supported by the I-CORE programme 'The Quantum Universe' of the Planning and Budgeting Committee and The Israel Science Foundation. A.G.-Y. acknowledges support by grants from the ISF, BSF, GIF, Minerva and FP7/ERC, and a Kimmel Investigator award. M.M.K. acknowledges support from Hubble and Carnegie-Princeton fellowships. E.O.O. acknowledges the Arye Dissentshik Career Development Chair and a grant from the Israeli MOST. J.C.W. is supported in part by the NSF. J.M.S. is supported by an NSF Postdoctoral Fellowship. A.V.F. acknowledges financial support from the TABASGO Foundation, the Richard and Rhoda Goldman Fund, the Christopher R. Redlich Fund and the NSF. The National Energy Research Scientific Computing Center, supported by the Office of Science of the US Department of Energy, provided staff, computational resources and data storage for this project. The Oskar Klein Centre is funded by the Swedish Research Council. We thank K. Clubb, O. Fox, P. Kelly, S. Tang and B. Sesar for their help with observations, and J. Groh, P. Crowther, M. Bersten, C. Fransson and E. Nakar for advice. Some of the data presented here were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. The observatory was made possible by the generous financial support of the W. M. Keck Foundation.

Author Contributions A.G.-Y. initiated the study, conducted analysis and wrote the manuscript. I.A. found the supernova, triggered rapid follow-up spectroscopy and contributed to light-curve analysis, observations, data reduction and manuscript preparation. E.O.O. contributed to analysis of early-time data, mass loss estimates, temperature evolution and manuscript preparation. S.B.-A. contributed to data reduction and to early light-curve and spectroscopic analysis. S.B.C. reduced Swift and Palomar 60-inch data, and contributed to spectroscopic reduction and analysis. M.M.K. provided APO data and contributed to manuscript preparation. Y.C. contributed to APO data reduction, early light-curve analysis and manuscript preparation. O.Y. contributed to observations and manuscript preparation. D.T. provided unpublished supernova light-curve templates and contributed to photometric analysis. J.M.S. provided spectroscopic reduction and advice, and contributed to HET spectroscopy. A.H. provided early Keck spectroscopy. A.D.C. contributed to observations and data reduction. F.T. reduced NOT data. J.S. provided NOT spectroscopy. D.P. provided Keck spectroscopy and conducted analysis. P.M.V. assisted with observations, spectroscopic analysis, and figure and manuscript preparation. P.E.N. is a PTF builder and contributed to the manuscript. S.R.K. is a PTF builder. A.V.F. provided Keck data and edited the manuscript. J.C.W. provided HET data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G.-Y. (avishay.gal-yam@weizmann.ac.il).

METHODS

Photometry. *r*-band images were obtained by the iPTF survey camera mounted on the Palomar 48-inch Schmidt telescope^{11,30}. Photometry was measured using our custom pipeline performing PSF photometry on iPTF images after removing a reference image constructed from pre-explosion data using image subtraction. Swift ultraviolet absolute AB magnitudes (Extended Data Fig. 1) were measured using standard pipeline reduction³¹ and were corrected for host-galaxy contamination using late-time Swift images.

Spectroscopy. Our earliest (15.5-h) and latest (69-d) spectra were obtained using the Deep Imaging Multi-Object Spectrograph (DEIMOS³²) spectrograph mounted on the Keck II 10-m telescope using the 600 line mm⁻¹ grating and an exposure time of 600 s. Additional spectra were obtained using ALFOSC mounted on the 2.56-m NOT telescope, LRS mounted on the 10.4-m HET telescope (day 3) and the Low Resolution Imaging Spectrometer³³ (LRIS) mounted on the Keck I 10-m telescope (day 6). All spectra were reduced using standard pipelines and are digitally available on WISEREP²⁹.

Spectroscopic observations are presented in Fig. 1. In Fig. 1a, the early spectrum of SN 2013cu reveals Wolf–Rayet wind features. The spectrum (black) is compared with WNL models⁷ (red and grey curves, offset vertically for clarity) showing remarkable similarity, both in line features (major species marked; strong He and N lines accompanied by Balmer lines indicate a WN6(h) classification) and in the continuum shape (demonstrated by overplotting the 56-kK model on the spectrum). The similarity in continuum shape to hot model spectra constrains any dust reddening to be minimal, indicating that any pre-existing circumstellar dust must have been destroyed; compare with the observed spectrum of the WN6(h) star HD 192163 (blue). Consistent with this conclusion, we detect no trace of Na D absorption lines. Figure 1b shows the emission line evolution during the first week. We replot the first spectrum (red), compared with spectra obtained ~3 d after explosion (using ALFOSC (blue) and LRS (cyan)) and ~6 d after explosion (magenta; additional details in A.G.-Y. *et al.*, manuscript in preparation). Spectra were scaled and offset for clarity with respect to the high-quality Keck/LRIS spectrum. Continuum shapes are identical (within calibration uncertainties) and consistent with the Rayleigh–Jeans tail of a hot Planck curve, indicating that the black-body peak remains in the ultraviolet at 6 d past explosion. By day 3, the initially strong Wolf–Rayet features disappear, with the exception of weaker He II lines and the C III/N III complex around 4,640 Å (marked), whereas the H α line remains constant at 3.4×10^{-15} erg s⁻¹ cm⁻² (with decreased equivalent width due to higher continuum level). The spectrum with a high signal-to-noise ratio on day 6 is almost featureless, except for weak H α emission (inset) with an intensity less than a tenth that of day 3, probably because the line-forming region has been cleared by the expanding ejecta. Figure 1c shows evidence that SN 2013cu is of type IIb. A spectrum of SN 2013cu 69 d after explosion (black) is compared with the prototypical type IIb supernovae SN 1993J at age 60 d (red) and SN 2011dh at age 43 d (green), and with the typical non-stripped type II-P supernova SN 2004et at age 45 d (blue). To allow for slight age differences and expansion-velocity variations, we align all spectra in wavelength using the weak blue-shifted Fe II lines at 4,924, 5,018 and 5,169 Å (marked with black vertical lines), because they are good tracers of the photosphere. Supernovae SN 2013cu, SN 2011dh and SN 1993J exhibit strong He I absorption at 5,876, 7,065 and 7,281 Å (marked with black vertical lines), whereas the weaker 6,678 Å absorption is similar in SN 2013cu and SN 2011dh. These He I lines are not detected in the type II-P supernova spectrum at all. However, SN 2013cu shows weaker Balmer absorption (marked with magenta vertical lines), and the H α absorption is not clearly defined. Both the spectral similarity to supernovae SN 1993J and SN 2011dh and the strong He I lines relative to H indicate that SN 2013cu is spectroscopically a type IIb supernova.

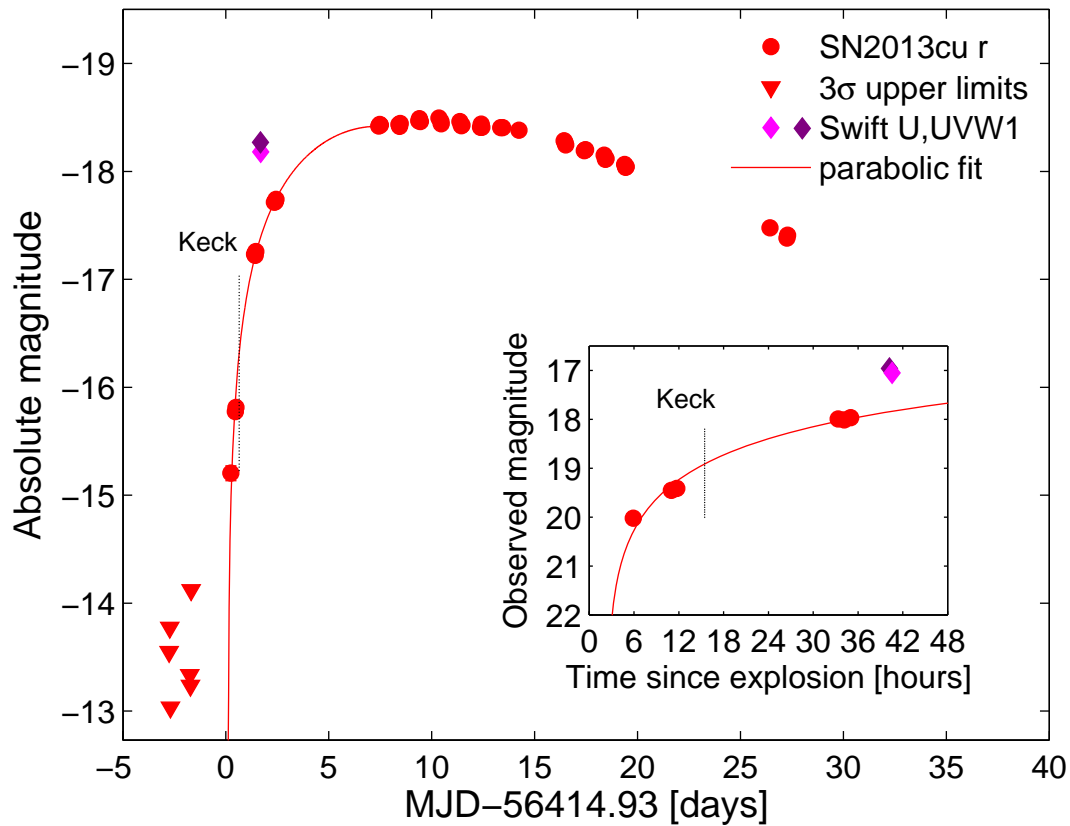
Flash spectroscopy. We define as ‘flash spectroscopy’ the technique of obtaining a set of spectroscopic data shortly enough after a supernova explosion that the observed spectrum is dominated by features related directly to the effects of the

shock breakout flash. In particular, flash-ionized CSM recombines and forms strong emission lines, revealing, for example, the elemental abundance and, thus, the Wolf–Rayet class of a supernova progenitor. In addition, emission line spectra provide a handle on the early temperature evolution, which is difficult to measure even using Swift ultraviolet photometry because the black-body peak is initially too far into the ultraviolet. This study provides strong motivation for future investigations using dedicated rapid-response spectrographs, such as FLOYDS^{3,34} and SEDM³⁵, responding to real-time triggers from high-cadence wide-field surveys¹².

Although Wolf–Rayet supernova progenitor stars are difficult to study using pre-explosion imaging (owing both to intrinsic low luminosity in the optical and infrared bands^{14,36} and to the possible confusing effect of a bright O/B companion³), we demonstrate that they are amenable to study using the flash spectroscopy method. WRSs belonging to the WN(h) class may have the most extensive winds¹⁴. Application of this method to WC and WO stars may require flash spectroscopy at even earlier epochs (~1 h after explosion), before the supernova ejecta sweep up the high-density wind. Reducing the latency between the explosion and the spectroscopy by an order of magnitude relative to our observations of SN 2013cu is possible using recently commissioned instrumentation (Extended Data Fig. 2).

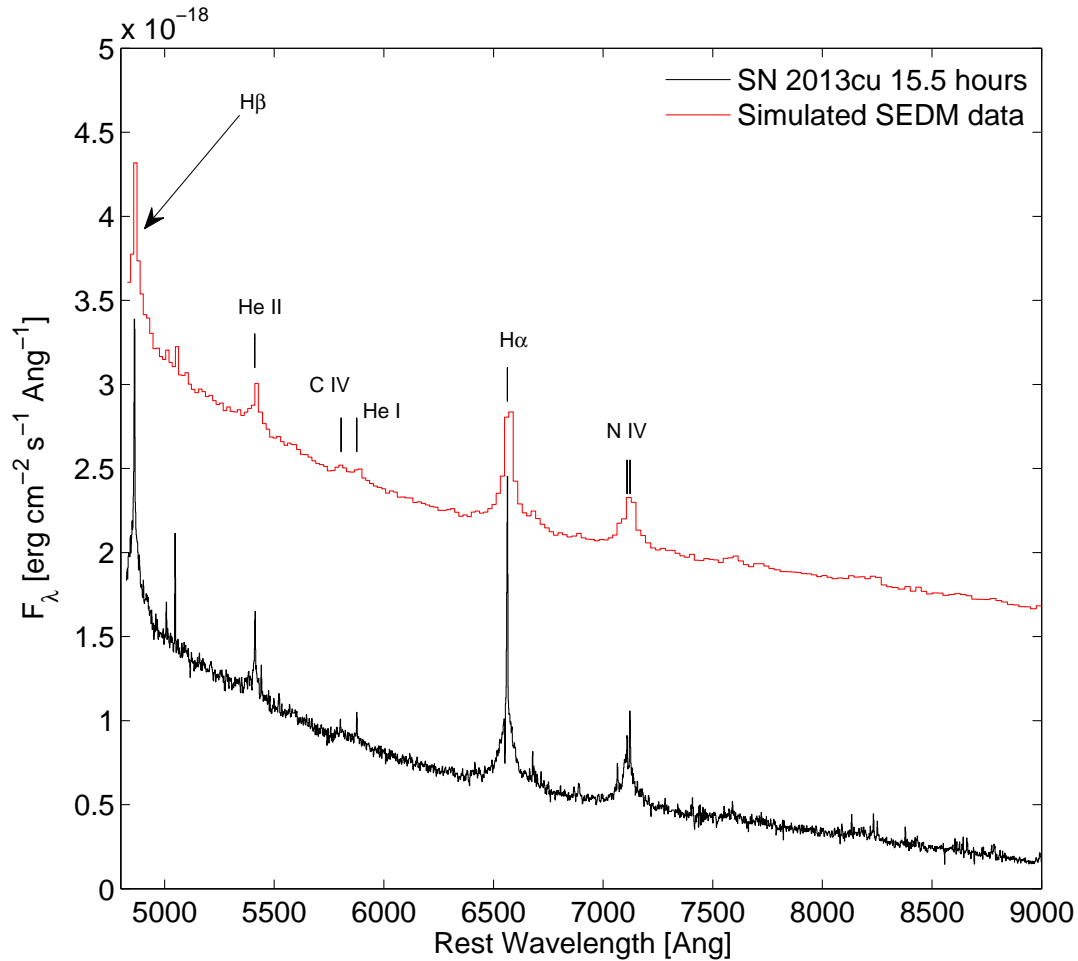
Unlike studies of supernova progenitors through pre-explosion imaging, the flash spectroscopy method can be applied to relatively distant objects (SN 2013cu is located 108 Mpc away, well beyond the 20-Mpc distance typical for pre-explosion studies) and to events in galaxies having no high-quality pre-explosion imaging, such as a large population of little-studied dwarf galaxies. Judging from local supernova rate measurements³⁷, ~300 explosions occur within 100 Mpc of Earth every year and can be potentially studied in this manner. The method thus allows routine spectroscopic studies of supernova progenitors that were previously only possible by extreme serendipity (for example for the progenitor of SN 1987A; ref. 38). Within a few years, there will be enough data for the flash spectroscopy method to be used to chart wind signatures from numerous supernova progenitors; in particular, it will be possible to study systematically the Wolf–Rayet progenitor population of stripped supernovae. We thus expect that this method will be broadly applied in the coming years.

30. Rau, A. *et al.* Exploring the optical transient sky with the Palomar Transient Factory. *Proc. Astron. Soc. Pacif.* **121**, 1334–1351 (2009).
31. Poole, T. S. *et al.* Photometric calibration of the Swift ultraviolet/optical telescope. *Mon. Not. R. Astron. Soc.* **383**, 627–645 (2008).
32. Faber, S. M. *et al.* The DEIMOS spectrograph for the Keck II Telescope: integration and testing. *Proc. SPIE* **4841**, 1657–1669 (2003).
33. Oke, J. B. *et al.* The Keck Low-Resolution Imaging Spectrometer. *Proc. Astron. Soc. Pacif.* **107**, 375–385 (1995).
34. Ben-Ami, S. *et al.* Discovery and early multi-wavelength measurements of the energetic type Ic supernova PTF12gzk: a massive-star explosion in a dwarf host galaxy. *Astrophys. J.* **760**, L33 (2012).
35. Ben-Ami, S. *et al.* The SED machine: a dedicated transient IFU spectrograph. *Proc. SPIE* **8446**, 844686 (2012).
36. Yoon, S.-C., Gräfener, G., Vink, J. S., Kozyreva, A. & Izzard, R. G. On the nature and detectability of type Ib/c supernova progenitors. *Astron. Astrophys.* **544**, L11 (2012).
37. Li, W. *et al.* Nearby supernova rates from the Lick Observatory Supernova Search - II. The observed luminosity functions and fractions of supernovae in a complete sample. *Mon. Not. R. Astron. Soc.* **412**, 1441–1472 (2011).
38. Walborn, N. R. *et al.* The spectrograms of Sanduleak –69°202, precursor to supernova 1987A in the Large Magellanic Cloud. *Astron. Astrophys.* **219**, 229–236 (1989).
39. Niemela, V. S., Ruiz, M. T. & Phillips, M. M. The supernova 1983k in NGC 4699 - clues to the nature of type II progenitors. *Astrophys. J.* **289**, 52–57 (1985).
40. Phillips, M. M. *et al.* The light curve of the plateau type II SN 1983K. *Proc. Astron. Soc. Pacif.* **102**, 299–305 (1990).



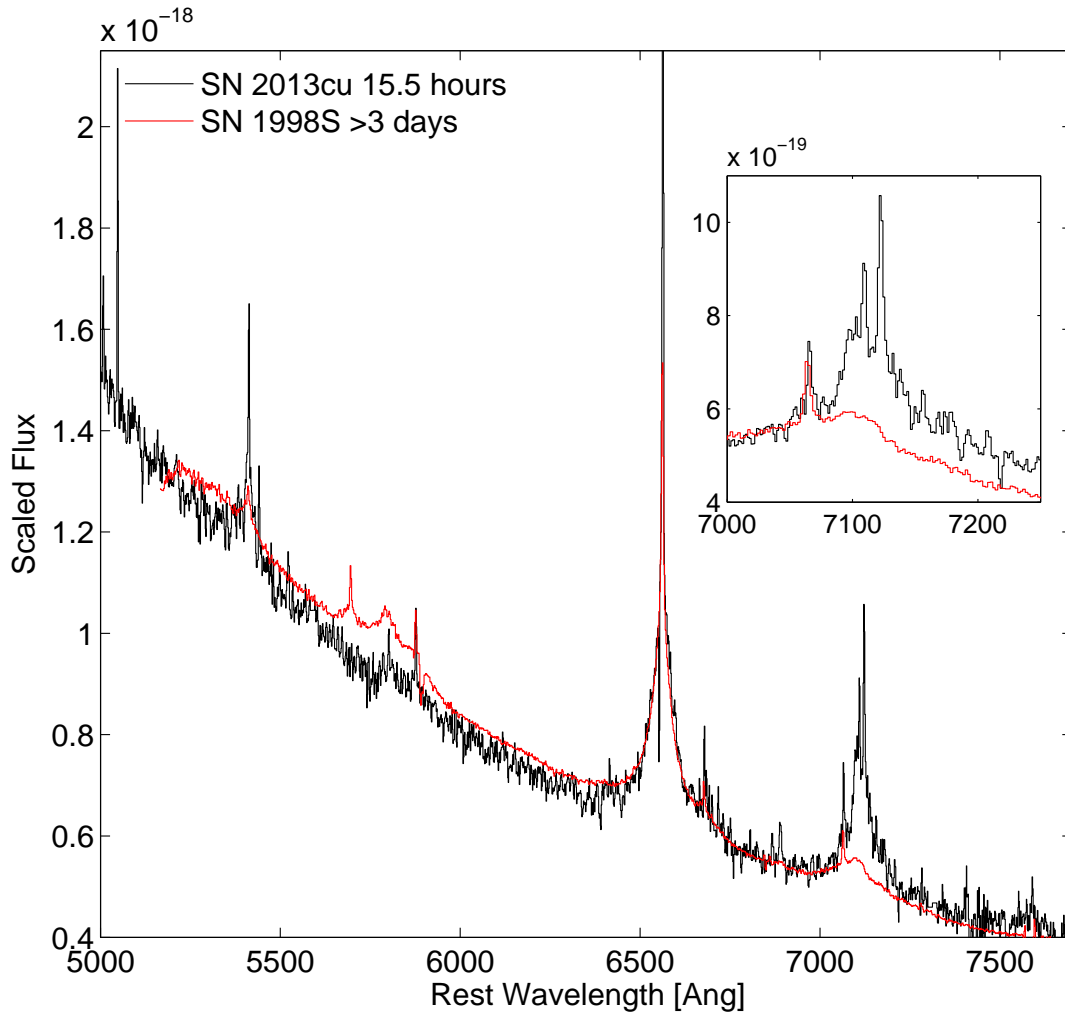
Extended Data Figure 1 | The r -band light curve of SN 2013cu. A parabolic model of the flux–time (red solid curve) describes the pre-peak data (1σ error bars) very well. Backward extrapolation indicates an explosion date of UTC 2013 May 2.93 ± 0.11 (MJD = 56414.93; 5.7 h before the first iPTF detection; see inset); we estimate the uncertainty from the scatter generated by

modifying the subset of points used in the fit. Our first Keck spectrum was obtained about 15.5 h after explosion (vertical dotted line). Early Swift ultraviolet photometry (diamonds) places a lower limit of $T = 25,000$ K on the black-body temperature measured 40 h after explosion.



Extended Data Figure 2 | Flash spectroscopy: rapid spectroscopic observations of supernovae during or shortly after shock breakout. This simulated SEDM spectrum (red) created by downgrading the observed Keck spectrum (black; resolution, $R = 2,000$) to the coarse SEDM resolution ($R = 100$) shows that the strong Wolf-Rayet lines (in this case the marked H, He and N lines) are still easily detectable and allow us to determine the Wolf-Rayet spectroscopic class. The SEDM is an IFU low-resolution spectrograph designed for robotic response to transient events, to be mounted on the Palomar 60-inch telescope almost continually. Responding to real-time triggers

from the iPTF survey operating on the same mountain, this instrument should be able to obtain low-resolution spectra within ~ 1 h of object detection. Because SEDM operates on a smaller telescope than Keck, SEDM data of similar quality to the simulated spectrum will require a relatively long integration. However, SEDM will be able to observe objects with much reduced latency, thus benefiting from stronger line intensities expected from the stronger shock breakout flash luminosity processed by a denser wind close to the progenitor star, potentially compensating for its reduced absolute sensitivity.



Extended Data Figure 3 | Comparison with early ‘Wolf-Rayet’ spectra of SN 1998S. Wolf-Rayet-like features similar to those we observed were previously noted in two cases, SN 1983K (refs 39,40) and SN 1998S (ref. 16), and persisted for many days after explosion. The spectra of SN 1983K, classified as a type II-P supernova, are unfortunately not available for comparison. Spectra of SN 1998S (type IIn) are shown here. The spectra have a similar blue continuum slope and a similar H α profile. The He II λ 5,411 and the N IV λ 7,115 complex are weaker in SN 1998S, and the strong lines of N and C (5,806 Å) are broad, consistent with an origin in shocked CSM¹⁶ rather than in an

undisturbed Wolf-Rayet wind. The inset shows a close-up view of the strong N IV λ 7,115 complex. N emission from SN 1998S is weak compared with He I λ 7,065 and shows a smooth, broad profile, whereas SN 2013cu exhibits a broad base (full-width at half-maximum, 2,500 km s⁻¹) as well as strong and narrow (unresolved) N IV lines. These observations are consistent with an origin for the N emission of SN 1998S in shocked (and perhaps N-rich) CSM¹⁶, whereas the narrow He I lines may come from a more distant, photo-ionized wind. In SN 2013cu even the narrow N lines are much stronger than He, indicating a wind with a Wolf-Rayet-like composition at all radii.

Practical quantum key distribution protocol without monitoring signal disturbance

Toshihiko Sasaki¹, Yoshihisa Yamamoto^{2,3} & Masato Koashi¹

Quantum cryptography^{1–8} exploits the fundamental laws of quantum mechanics to provide a secure way to exchange private information. Such an exchange requires a common random bit sequence, called a key, to be shared secretly between the sender and the receiver. The basic idea behind quantum key distribution (QKD) has widely been understood as the property that any attempt to distinguish encoded quantum states causes a disturbance in the signal. As a result, implementation of a QKD protocol involves an estimation of the experimental parameters influenced by the eavesdropper's intervention, which is achieved by randomly sampling the signal. If the estimation of many parameters with high precision is required, the portion of the signal that is sacrificed increases, thus decreasing the efficiency of the protocol^{9,10}. Here we propose a QKD protocol based on an entirely different principle. The sender encodes a bit sequence onto non-orthogonal quantum states and the receiver randomly dictates how a single bit should be calculated from the sequence. The eavesdropper, who is unable to learn the whole of the sequence, cannot guess the bit value correctly. An achievable rate of secure key distribution is calculated by considering complementary choices between quantum measurements of two conjugate observables¹¹. We found that a practical implementation using a laser pulse train achieves a key rate comparable to a decoy-state QKD protocol^{12–14}, an often-used technique for lasers. It also has a better tolerance of bit errors and of finite-sized-key effects. We anticipate that this finding will give new insight into how the probabilistic nature of quantum mechanics can be related to secure communication, and will facilitate the simple and efficient use of conventional lasers for QKD.

In a QKD protocol, the sender Alice and the receiver Bob repeat transmission of quantum signals and accumulate raw bits of data through quantum measurements. Using public communication, each of them discards the apparently useless portion of the raw data to form a sifted key. The sifted key of length N is then processed into the final key of a shorter length through error reconciliation and privacy amplification. Denoting the costs of the two procedures as H_{ER} and H_{PA} , the net production length G of the secure final key is given by

$$G = N(1 - H_{\text{ER}} - H_{\text{PA}}) \quad (1)$$

When the bit errors between Alice's and Bob's sifted key occur at a rate e_{bit} , the ideal cost of error reconciliation in the asymptotic limit of large N is given by $H_{\text{ER}} = h(e_{\text{bit}})$ as Shannon entropy, with $h(x) = -x \log_2 x - (1-x) \log_2 (1-x)$. The cost H_{PA} depends on how much information on the sifted key has leaked to an eavesdropper Eve. For example, it is given by $H_{\text{PA}} = h(e_{\text{bit}})$ in a simple proof¹⁵ for the Bennett–Brassard 1984 (BB84) protocol¹. In general, the formula varies depending on protocols and security proofs, and parameters other than e_{bit} are often monitored in the protocol and enter into the formula of H_{PA} . Nevertheless, so far H_{PA} has always been an increasing function of the amount of disturbance. This implies that the conventional QKD protocols^{1–8,12–14} inherently rely on the original version of Heisenberg's uncertainty principle, which dictates that the more information Eve has obtained, the more disturbance she should have caused on the signal.

What we propose here is an entirely new approach to establishing private correlations between Alice and Bob under the presence of an eavesdropper Eve, in which the leaked information to Eve is bounded regardless of the disturbance that she causes on the quantum signal. The main idea is to encode many raw key bits on quantum systems coherently such that only a few bits can be read out at the same time, which enables Bob to specify randomly how the sifted key bit is calculated from the raw key bits. This randomness makes it hard for Eve to guess the calculated bit from what little knowledge on the raw key bits she has acquired.

Let us explain our QKD protocol in more detail using the schematics shown in Fig. 1a. The protocol proceeds as follows. (I) Alice encodes a random L -bit sequence $s_1 s_2 \dots s_L$ on a weak signal. For the understanding of the basic idea, here we assume that the encoded signal is a single-photon state of L optical pulses

$$|\Psi_1\rangle = \frac{1}{\sqrt{L}} \sum_{k=1}^L (-1)^{s_k} |k\rangle \quad (2)$$

where the photon is in the k th pulse for state $|k\rangle$. (II) After possible intervention by Eve, Bob receives the signal. (III) An independent random

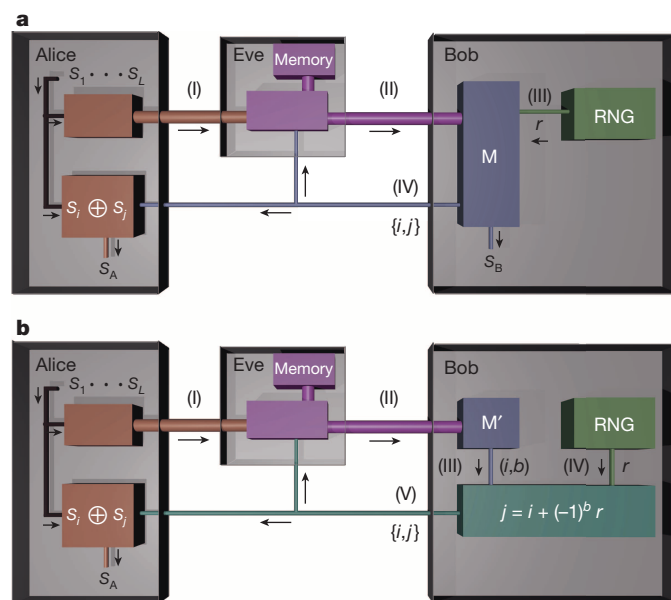


Figure 1 | Basic idea behind the proposed QKD scheme. **a**, **b**, Quantum signals flow through thick lines and classical ones through thin lines, in the order indicated by the Roman numerals. Eve tries to guess Alice's bit $s_A = s_i \oplus s_j$ in both figures, where indices $\{i, j\}$ are announced by Bob. In **a**, Bob conducts measurement M following random number generator RNG to guess s_A . In **b**, Bob conducts measurement M' prior to RNG, making it hard for anyone to guess s_A . As the procedures to generate indices $\{i, j\}$ in both figures are identical, every strategy used by Eve in **a** should work equally well in **b**.

¹Photon Science Center, Graduate School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. ²E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA.

³National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.

number generator (RNG) announces a random value $r \in \{1, \dots, L-1\}$. (IV) Through an optical interference measurement M , Bob tries to determine the value of $s_i \oplus s_j$ for a pair of indices $\{i, j\} \subset \{1, \dots, L\}$ satisfying $j - i = \pm r \pmod{L}$. Here the symbol \oplus denotes summation modulo 2. In measurement M , Bob splits each pulse by a half beam-splitter and then superposes the k th and the k' th half pulses ($k' = k + r \pmod{L}$, $k = 1, \dots, L$) to measure the phase difference by detecting a photon. Whenever a photon is detected from the superposed i th and j th pulses, Bob announces $\{i, j\}$ and records the measured phase difference as his sifted key bit s_B . Alice records $s_A = s_i \oplus s_j$ as her sifted key bit. As shown in the Methods, if Bob receives the state $|\Psi_1\rangle$ intact, he learns s_A without errors.

We are now interested in how well Eve can guess the value of s_A . Figure 1a alone is not conclusive in this regard, because she has a control over the decision process of the indices $\{i, j\}$ through feeding a modified signal to Bob at step (II). To show that Eve's control is quite limited, consider another measurement procedure by Bob shown in Fig. 1b. In measurement M' , Bob simply measures the location of the photon in the incoming L pulses to determine one of the indices, i . He also generates a random bit b . Subsequently, the RNG announces r , which determines the other index as $j = i + (-1)^b r \pmod{L}$. As is proved in the Methods, this procedure is equivalent to M as far as the production of outcome $\{i, j\}$ is concerned. Hence, it suffices to show Eve's ignorance of s_A in Fig. 1b.

An intuitive reasoning for the ignorance is given as follows. Alice has emitted just one photon, so most of the L bits should be unknown to Eve when $L \gg 1$. In Fig. 1b, Eve's intervention affects only the decision of index i , and the other index j is chosen randomly from the rest of the $L-1$ bits through the random number r . We may thus expect that Eve has little information on s_j , and hence on $s_A = s_i \oplus s_j$.

What is remarkable here is that the above argument has no reference to how much Eve has disturbed the signal received by Bob. To make a rigorous security proof, we have only to show that Alice's sifted key bit s_A in Fig. 1b can be accumulated and converted to a secure final key, based on the fact that r is random and independent of i . There is no need to mention directly the state fed to Bob by Eve in the proof, and it is still valid for any attack strategy by Eve.

The difficulty in guessing the value of randomly chosen bit s_j appearing in Fig. 1b has been discussed in a slightly different context and called the information causality¹⁶. Our QKD scheme may be regarded as the combination of the information causality, which holds for classical and quantum signals alike, and the complementarity, which is unique to quantum mechanics. Bob's measurement in Fig. 1a reveals the phase difference, a wave-like property, while that in Fig. 1b identifies the location of the photon, a particle-like property. In quantum mechanics, such different measurements may result in incompatible consequences. In fact, Bob learns s_A in Fig. 1a, whereas in Fig. 1b the information causality forbids anyone from learning s_A , including Bob. The mere possibility of Bob's choosing the latter prevents Eve from learning s_A even if Bob has actually chosen the former.

It is also worth mentioning how our QKD protocol differs from the B92 protocol³. They are similar at many points. In both, Alice encodes the bit values on non-orthogonal states. Bob dictates which of the bits should be used. The dictation may be tampered with by Eve via modification of the signals, because it is based on the outcomes of Bob's measurement on them. It is, however, only our protocol that has a complementary scenario shown in Fig. 1b, which substantiates the existence of inherent randomness in Bob's dictation, beyond the reach of Eve's tampering.

The above basic idea can be implemented simply by a weak coherent laser pulse train as a light source and a variable-delay interferometer at the receiver (Fig. 2), which we name the round-robin differential phase-shift (RRDPS) QKD protocol. The setup is exactly the same as the differential phase-shift QKD protocol^{7,17,18} except that the fixed delay line in the original is replaced by a variable delay line. For a security proof, we adopt a simple characterization of the source about the total photon number ν in the L -pulse train stated in the form of

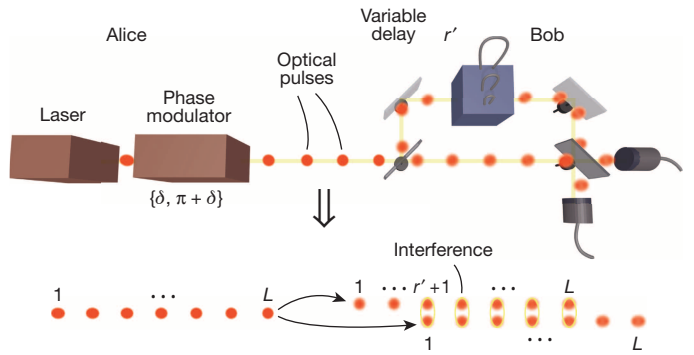


Figure 2 | Practical implementation of the proposed QKD scheme. Alice's laser emits a train of L pulses with interval T . She applies phase shift $\{0, \pi\}$ on each pulse according to a random bit sequence $s_1 \dots s_L$. Bob splits the received train into two beams and superposes them after a random delay $r'T$ ($r' \in \{1, \dots, L-1\}$). Detection of a photon determines Bob's sifted key bit s_B , and he announces the indices $\{i, j\}$ of the corresponding pair of pulses. Alice adopts $s_A = s_i \oplus s_j$ as her sifted key bit. The key rate is improved by applying random phase δ on each train (see the Methods).

$$\Pr(\nu > \nu_{\text{th}}) \leq e_{\text{src}} \quad (3)$$

with an integer $\nu_{\text{th}} < \frac{L-1}{2}$ and a constant e_{src} . Let Q be the empirical rate of detection $Q = N/N_{\text{em}}$ when a sifted key of length N is generated through N_{em} rounds of transmitting L -pulse trains. Then we can derive an asymptotic formula for the net production length of the secure key (see the Methods)

$$G = N \left[1 - h(e_{\text{bit}}) - \frac{e_{\text{src}}}{Q} - \left(1 - \frac{e_{\text{src}}}{Q} \right) h\left(\frac{\nu_{\text{th}}}{L-1}\right) \right] \quad (4)$$

For clarity, let us consider the case where a nonclassical light source with $e_{\text{src}} = 0$ is used instead of the laser, for which

$$G = N \left[1 - h(e_{\text{bit}}) - h\left(\frac{\nu_{\text{th}}}{L-1}\right) \right] \quad (5)$$

The case with $\nu_{\text{th}} = 1$ corresponds to the state $|\Psi_1\rangle$ used in the explanation of the basic idea. In equation (5), the third term $H_{\text{PA}} = h(\nu_{\text{th}}/(L-1))$ is a constant, which is in stark contrast with the conventional QKD protocols for which H_{PA} depends on the disturbance. The constant value of H_{PA} leads to two advantages of the new QKD protocol: (1) It has high tolerance of bit errors. For example, for $L = 128$ and $\nu_{\text{th}} = 1$, G is positive up to $e_{\text{bit}} = 0.35$. There is no fundamental limit on the error threshold smaller than 50%. (2) The secrecy of the final key is established after shortening the key length via privacy amplification by a fixed and predetermined fraction H_{PA} . There is no need to sacrifice a randomly chosen subset of signals to estimate an appropriate value of H_{PA} , which affects the rate of finite-sized key generation^{9,10,19–21}.

For the use of weak coherent pulses (WCPs) from a conventional laser, we show examples of asymptotic key rates per pulse, $G/(LN_{\text{em}})$, as a function of channel transmission η in Fig. 3. For ground-based transmission, an optical fibre of 50 km decreases η by a factor of 10, whereas $\eta = 10^{-4}$ to 10^{-5} is expected for satellite-based transmission^{22–24}. Figure 3 also shows rates of the BB84 protocol for comparison. When WCPs with the second-order correlation $g^{(2)}(0) = 1$ or a realistic single-photon source^{25–27} with $g^{(2)}(0) = 0.01$ are used for BB84, the multi-photon emission from the source is exploited by Eve via photon-number splitting attacks^{28,29}, resulting in a poor key-rate scaling of $O(\eta^2)$. The present scheme with WCPs has a better scaling, close to $O(\eta)$, and surpasses the WCP-based or the single-photon-based BB84 protocol for small η .

There is a popular technique called decoy-state QKD^{12–14}, in which pulses with different amplitudes are randomly mixed in the signal to monitor the photon-number-splitting attacks. It stands in sharp contrast to our protocol. Figure 3 also shows the asymptotic key rate for the ideal decoy-state BB84 protocol¹³, in which the statistics of single-photon

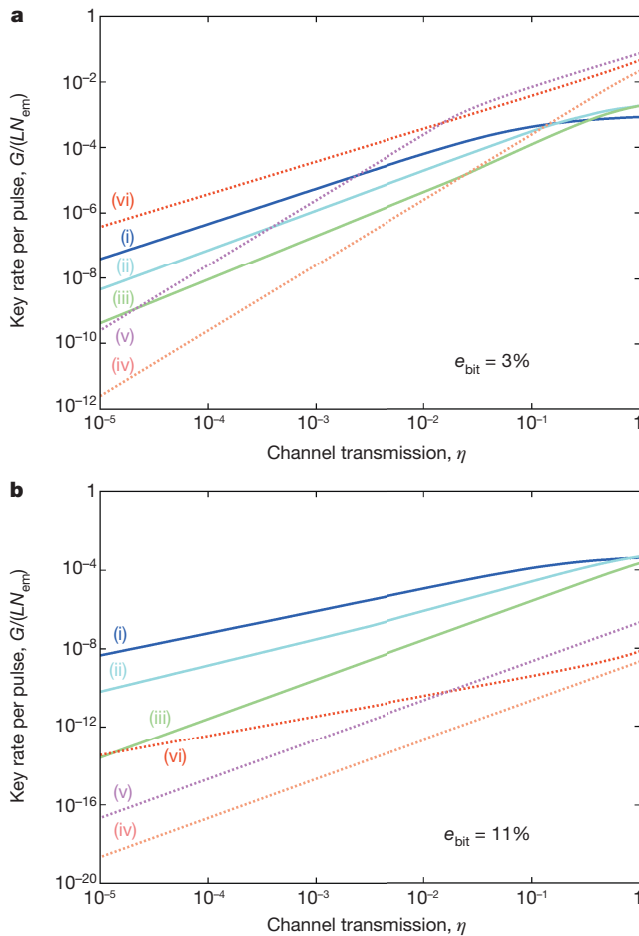


Figure 3 | Key rates versus channel transmission. **a**, The rates for $e_{\text{bit}} = 0.03$. **b**, The rates for $e_{\text{bit}} = 0.11$. Lines labelled (i)–(iii) represent the proposed protocol with $L = 128, 32$ and 16 . The rates are optimized over the choice of v_{th} and the mean photon number μ of a WCP through the relation $Q = L\eta\mu e^{-L\eta\mu}/2$ and $e_{\text{src}} = 1 - \sum_{v=0}^{v_{\text{th}}} \mu^{-v}/v!$. The optimized value of μ is around 0.05 for line (i) when $\eta < 0.01$. Lines labelled (iv)–(vi) represent BB84 protocols with double-pulse phase coding, using WCPs (iv), realistic single-photon source with the second-order correlation $g^{(2)}(0) = 0.01$ (v), and WCPs with infinite decoy states (vi).

emission events are precisely characterized via decoy signals. Although the asymptotic rate is better than our protocol by one order of magnitude, for a finite-sized key the decoy-state BB84 protocol suffers from a trade-off between the overhead of processing a large-sized key and the inefficiency of inserting many decoy signals to reach a required accuracy in the estimation of parameters^{9,10}. Our protocol is much simpler in this regard, requiring no sampling for determining H_{PA} . As a result, a positive key rate is achieved even with N being as small as 10^3 (see Methods). For a higher bit-error rate, our protocol becomes better than the decoy-state BB84 even in the asymptotic limit, owing to its high tolerance on the errors. If we consider the use of modern digital coherent communication systems with 40 Gbits s^{-1} differential phase-shift signals and assume the receiver's overall detection efficiency to be 10%, we can generate a secure key at a rate of 200 bits s^{-1} for a channel length of 200 km and an error rate of 11%.

The variable delay used in our scheme will be implemented as a series of switchable optical delay lines of $T, 2T, 4T, 8T, \dots$, where T is the time interval between the neighbouring pulses. Because the delay is fixed for each train of L pulses, the switching speed can be much slower than T , and it affects only the duty ratio.

The proposed QKD protocol demonstrates that spreading quantum information coherently over hundreds of quantum systems such as optical

pulses provides a novel way of utilizing it for secure communication. The fact that the quantum effect survives under large noise suggests that similar encoding techniques may be useful for other applications of quantum information working in the presence of noise.

METHODS SUMMARY

Bob's alternative choices of measurements. Let $+_L$ denote summation modulo L . When a single-photon input state $\hat{\rho}$ is fed to measurement M , Bob announces $\{k, k+_L\}$ and obtains $s_B = s$ at probability $\langle k, s | \hat{\rho} | k, s \rangle / 2$ with $|k, s\rangle = (|k\rangle + (-1)^s |k+_L\rangle) / \sqrt{2}$. Given that $\langle k, s | \Psi_1 \rangle = 0$ when $s \neq s_k \oplus s_{k+_L}$, Bob's guess s_B is always equal to s_A if he has received state $|\Psi_1\rangle$. The probability of announcing $\{i, j\}$ (note that $\{j, i\}$ is regarded as the same value) is calculated to be $P(\{i, j\}) = [P(i) + P(j)] [\delta_{i+_L, r, j} + \delta_{j+_L, i, r}] / 2$, where $P(k) = \langle k | \hat{\rho} | k \rangle$ is the probability of finding a photon in the k th pulse.

The calculation of $P(\{i, j\})$ for the case of Fig. 1b also leads to the same expression. This shows that the relation between the quantum signal received from Eve and the announced value $\{i, j\}$ is identical for Fig. 1a and Fig. 1b.

Derivation of secure key rates. The random phase shift δ enables Alice to tag each of the rounds with $v > v_{\text{th}}$ in principle³⁰. We assume that this tagged portion, at most $N e_{\text{src}} / Q$ bits, is fully leaked to Eve, leading to the $-e_{\text{src}} / Q$ term in equation (4).

For the untagged portion, it can be shown that the sequence $s_1 s_2 \dots s_L$ is equivalent to the outcome of $\{|0\rangle, |1\rangle\}$ -basis measurement on L qubits prepared in a state fulfilling the promise that, if they are measured in a conjugate $\{|+\rangle, |-\rangle\}$ -basis, no more than v_{th} qubits are found to be in the $|-\rangle$ state. The key bit $s_A = s_i \oplus s_j$ is then given by $\{|0\rangle, |1\rangle\}$ -basis measurement on qubit j after a controlled-NOT operation on qubits i and j . It can be shown that the probability of finding qubit j in the state $|-\rangle$ is at most $v_{\text{th}} / (L - 1)$, leading¹¹ to the remaining term in equation (4).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 November 2013; accepted 27 March 2014.

- Bennett, C. H. & Brassard, G. In *Proc. IEEE Int. Conf. on Computers, Systems and Signal Processing* 175–179 (IEEE Press, 1984).
- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Bennett, C. H. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.* **68**, 3121–3124 (1992).
- Bruß, D. Optimal eavesdropping in quantum cryptography with six states. *Phys. Rev. Lett.* **81**, 3018–3021 (1998).
- Scarani, V., Acín, A., Ribordy, G. & Gisin, N. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Phys. Rev. Lett.* **92**, 057901 (2004).
- Stucki, D., Brunner, N., Gisin, N., Scarani, V. & Zbinden, H. Fast and simple one-way quantum key distribution. *Appl. Phys. Lett.* **87**, 194108 (2005).
- Inoue, K., Waks, E. & Yamamoto, Y. Differential-phase-shift quantum key distribution using coherent light. *Phys. Rev. A* **68**, 022317 (2003).
- Grosshans, F. & Grangier, P. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.* **88**, 057902 (2002).
- Cai, R. Y. & Scarani, V. Finite-key analysis for practical implementations of quantum key distribution. *New J. Phys.* **11**, 045024 (2009).
- Hayashi, M. & Nakayama, R. Security analysis of the decoy method with the Bennett-Brassard 1984 protocol for finite key lengths. Preprint at <http://arxiv.org/abs/1302.4139> (2013).
- Koashi, M. Simple security proof of quantum key distribution based on complementarity. *New J. Phys.* **11**, 045018 (2009).
- Hwang, W.-Y. Quantum key distribution with high loss: toward global secure communication. *Phys. Rev. Lett.* **91**, 057901 (2003).
- Lo, H.-K., Ma, X. & Chen, K. Decoy state quantum key distribution. *Phys. Rev. Lett.* **94**, 230504 (2005).
- Wang, X.-B. Beating the photon-number-splitting attack in practical quantum cryptography. *Phys. Rev. Lett.* **94**, 230503 (2005).
- Shor, P. W. & Preskill, J. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85**, 441–444 (2000).
- Pawlowski, M. et al. Information causality as a physical principle. *Nature* **461**, 1101–1104 (2009).
- Takesue, H. et al. Quantum key distribution over a 40-dB channel loss using superconducting single-photon detectors. *Nature Photon.* **1**, 343–348 (2007).
- Tamaki, K., Koashi, M. & Kato, G. Unconditional security of coherent-state-based differential phase shift quantum key distribution protocol with block-wise phase randomization. Preprint at <http://arxiv.org/abs/1208.1995> (2012).
- Mayers, D. Unconditional security in quantum cryptography. *J. ACM* **48**, 351–406 (2001).
- Hayashi, M. & Tsurumaru, T. Concise and tight security analysis of the Bennett-Brassard 1984 protocol with finite key lengths. *New J. Phys.* **14**, 093014 (2012).
- Tomamichel, M., Lim, C. C. W., Gisin, N. & Renner, R. Tight finite-key analysis for quantum cryptography. *Nature Commun.* **3**, 634 (2012).

22. Bourgoin, J. *et al.* A comprehensive design and performance analysis of low earth orbit satellite quantum communication. *New J. Phys.* **15**, 023006 (2013).
23. Nauerth, S. *et al.* Air-to-ground quantum communication. *Nature Photon.* **7**, 382–386 (2013).
24. Wang, J.-Y. *et al.* Direct and full-scale experimental verifications towards ground-satellite quantum key distribution. *Nature Photon.* **7**, 387–393 (2013).
25. He, Y.-M. *et al.* On-demand semiconductor single-photon source with near-unity indistinguishability. *Nature Nanotechnol.* **8**, 213–217 (2013).
26. Yuan, Z. *et al.* Electrically driven single-photon source. *Science* **295**, 102–105 (2002).
27. Claudon, J. *et al.* A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photon.* **4**, 174–177 (2010).
28. Huttner, B., Imoto, N., Gisin, N. & Mor, T. Quantum cryptography with coherent states. *Phys. Rev. A* **51**, 1863–1869 (1995).
29. Brassard, G., Lütkenhaus, N., Mor, T. & Sanders, B. C. Limitations on practical quantum cryptography. *Phys. Rev. Lett.* **85**, 1330–1333 (2000).
30. Gottesman, D., Lo, H.-K., Lütkenhaus, N. & Preskill, J. Security of quantum key distribution with imperfect device. *Quant. Inf. Comput.* **4**, 325 (2004).

Acknowledgements We thank H. Takesue and K. Tamaki for helpful discussions. This work was supported by the Funding Program for World-Leading Innovative R & D on Science and Technology (FIRST), Grant-in-Aid for Scientific Research on Innovative Areas number 21102008 (MEXT), and Photon Frontier Network Program (MEXT).

Author Contributions All authors contributed to the initial conception of the ideas, to the working out of details, and to the writing and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. (koashi@qi.tu-tokyo.ac.jp).

METHODS

Bob's alternative choices of measurements. Let $+_L$ denote summation modulo L , and $\hat{P}(|\phi\rangle) := |\phi\rangle\langle\phi|$. Bob's measurement M is fully characterized by a set of operators

$$\hat{E}_{k,s}^{(r)} := \frac{1}{2} \hat{P} \left(\frac{|k\rangle + (-1)^s |k+_L r\rangle}{\sqrt{2}} \right) \quad (6)$$

where the probability of outcome (k, s) ($k \in \{1, \dots, L\}, s \in \{0, 1\}$) is given by $\text{Tr}(\hat{\rho} \hat{E}_{k,s}^{(r)})$ for single-photon input state $\hat{\rho}$. From this outcome, he announces $\{k, k+_L r\}$ (the order being irrelevant) and adopts $s_B = s$. Given that $\langle \Psi_1 | \hat{E}_{k,s}^{(r)} | \Psi_1 \rangle = 0$ when $s \neq s_k \oplus s_{k+_L r}$, Bob's guess s_B is always equal to s_A if he has received state $|\Psi_1\rangle$.

In measurement M , the probability $P(\{i, j\})$ of announcing $\{i, j\}$ (where $i \neq j$) is given by $\sum_s \text{Tr}(\hat{\rho} \hat{E}_{i,s}^{(r)}) \delta_{i+_L r, j} + \sum_s \text{Tr}(\hat{\rho} \hat{E}_{j,s}^{(r)}) \delta_{j+_L r, i}$, where $\delta_{x,y}$ is 1 for $x = y$ and 0 for $x \neq y$. This is calculated to be

$$P(\{i, j\}) = [P(i) + P(j)] [\delta_{i+_L r, j} + \delta_{j+_L r, i}] / 2 \quad (7)$$

where $P(k) = \langle k | \hat{\rho} | k \rangle$ is the probability of finding a photon in the k th pulse.

In Fig. 1b, Bob announces $\{k, k+_L (-1)^b r\}$ ($k = 1, \dots, L; b = 0, 1$) at probability $P(k)/2$. Noticing that $P(\{i, j\})$ is contributed from the cases $k = i$ and $k = j$, we see that it is also given by equation (7). This shows that the relation between the quantum signal received from Eve and the announced value $\{i, j\}$ is identical for Fig. 1a and Fig. 1b. Therefore, if a statement regarding Eve's knowledge about Alice's sifted key is proved for Fig. 1b, it should also be true for Fig. 1a.

Derivation of secure key rates. Here we give a security proof and derive the final key rate for the proposed QKD protocol shown in Fig. 2. We first show that Alice's random bit sequence $s_1 \dots s_L$ can be regarded as an outcome of Z-basis measurement on L qubits. Let $|\Psi\rangle$ be the state of an L -pulse train emitted from the laser source. In the actual setup, Alice chooses $s_1 \dots s_L$ randomly and applies phase shifts accordingly, resulting in the emitted state $\otimes_k (-1)^{s_k \hat{n}_k} |\Psi\rangle$, where $\hat{n}_k := \hat{a}_k^\dagger \hat{a}_k$ is the photon-number operator for the k th pulse. Instead¹⁸, she could prepare L qubits and the L pulses in an entangled state

$$2^{-L/2} \bigotimes_{k=1}^L \sum_{s_k=0,1} |s_k\rangle_k (-1)^{s_k \hat{n}_k} |\Psi\rangle \quad (8)$$

where $\{|0\rangle_k, |1\rangle_k\}$ is the Z-basis states of the k th qubit. The states of the L pulses are identical to those in the actual setup, and if Alice needs bit value s_k , she may simply measure the k th qubit on the Z basis.

It is useful for later discussion to ask what happens if Alice measures the L qubits in the X basis $\{|+\rangle, |-\rangle\}$ with $|\pm\rangle := 2^{-1/2}(|0\rangle \pm |1\rangle)$. Let n_- be the number of qubits found in state $|-\rangle$. The statistics of n_- is related to the photon number distribution in $|\Psi\rangle$. In fact, it is seen from equation (8) that if the k th pulse contains an even number of photons, the state of the k th qubit is $|+\rangle_k$, and if the number is odd, the state is $|-\rangle_k$. Hence n_- is no larger than the total photon number. The argument so far holds for any pure state $|\Psi\rangle$, so it is also true when the source emits a mixed state. We thus conclude that, if the source fulfils equation (3)

$$\Pr(n_- > v_{\text{th}}) \leq e_{\text{src}} \quad (9)$$

Next, we relate Bob's apparatus in Fig. 2 to measurement M . We assume that the detectors can discriminate between a single photon from two or more photons, and that dark countings and inefficiency can be equivalently ascribed to a property of the transmission channel. Bob declares successful detection when a photon is detected from a superposed pulse and no other detection occurs in the whole pulse train. This ensures that the detected signal comes from a single-photon state. When the delay is $r'T$, the measurement is characterized similarly to equation (6) by operators

$$\hat{F}_{k',s}^{(r')} := \frac{1}{2} \hat{P} \left(\frac{|k'\rangle + (-1)^s |k'+_L r'\rangle}{\sqrt{2}} \right) \quad (10)$$

except that it is defined only if $1 \leq k' \leq L - r'$. To see that this is equivalent to measurement M except at an efficiency of 1/2, introduce an auxiliary random bit c to define $r = r'$ and $k = k'$ if $c = 0$, while $r = L - r'$ and $k = k' + r'$ if $c = 1$. Then r is uniformly random. Given r , the probability of outcome (k, s) for input state $\hat{\rho}$ is written as

$$\Pr(k, s) = \frac{1}{2} \text{Tr}(\hat{\rho} \hat{E}_{k,s}^{(r)}) + \frac{1}{2} \text{Tr}(\hat{\rho} \hat{E}_{k+_L r-L, s}^{(r)}) \quad (11)$$

where it is understood that $\hat{E}_{k',s}^{(r')} = 0$ for $k' \geq L - r' + 1$ or $k' \leq 0$. It turns out that one of the terms always vanishes and $\Pr(k, s) = \text{Tr}(\hat{\rho} \hat{E}_{k,s}^{(r)})/2$. We thus conclude that Bob's apparatus is equivalent to measurement M preceded by a filter that allows only single-photon states to pass through with efficiency 1/2.

To assess how much Eve knows about Alice's key bit s_A , we may assume that Bob carries out measurement M' . Learning i, b and r , Alice applies a controlled-NOT operation to the qubit i as control and qubit $j = i + (-1)^b r \pmod{L}$ as target. The key bit $s_A = s_i \oplus s_j$ in the original protocol is now equivalent to the outcome of Z-basis measurement on qubit j . If one measures this qubit in X basis instead, the probability of a 'phase error', namely, of finding it in the $|-\rangle$ state, is no more than e_{ph} , defined by

$$e_{\text{ph}} = \frac{e_{\text{src}}}{Q} + \left(1 - \frac{e_{\text{src}}}{Q}\right) \frac{v_{\text{th}}}{L-1} \quad (12)$$

This is because the controlled-NOT operation does not affect the X eigenstates of the target, the index j is chosen uniformly from all qubits except the i th via random number r , and finally equation (9) ensures that among the N rounds contributing the sifted key, at least $N - N_{\text{em}} e_{\text{src}}$ rounds satisfy $n_- \leq v_{\text{th}}$ in the limit of large N . Then, if $e_{\text{ph}} < 1/2$, she can extract a secure final key of length $N[1 - h(e_{\text{ph}})]$ by privacy amplification¹¹. Bob composes his sifted key from bit s_B in each round. The error reconciliation will be achieved by letting Alice send $Nh(e_{\text{bit}})$ bits of encrypted information to Bob such that he can reconcile his sifted key to Alice's. The net production length is then given by

$$G = N[1 - h(e_{\text{bit}}) - h(e_{\text{ph}})] \quad (13)$$

This rate can be improved by applying a common random optical phase shift δ to all the L pulses in the actual protocol. This makes the emitted quantum state of the train into a classical mixture of states with fixed total photon numbers, enabling Alice to tag each of the rounds with $v > v_{\text{th}}$ in principle though she need not do so in practice³⁰. We may then assume that Eve completely knows the sifted key bits for the tagged portion (at most N_{src}/Q bits), while the rest is treated as if $e_{\text{src}} = 0$. This leads to equation (4) in the main text and is used in Fig. 3.

If we omit the random optical phase shift δ , the rate for $L = 128$ and $e_{\text{bit}} = 0.03$ decreases by about 10% from the rate shown in Fig. 3. On the other hand, if we are allowed to assume that the emitted photon number obeys a Poissonian distribution, the rate for $L = 128$ and $e_{\text{bit}} = 0.03$ is larger than the one shown in Fig. 3 by about 30% even if we omit the random optical phase shift.

Finally, we briefly discuss an expected behaviour of our protocol for a finite-sized key. Let $\bar{f}(k; n, p) := \sum_{j \geq k} p^j (1-p)^{n-j} n! / [j!(n-j)!]$ be the tail distribution for finding more than k successful events in a binomial distribution. Except for a probability $\epsilon_1 := \bar{f}(Nr_1; N_{\text{em}}, e_{\text{src}})$, we may choose Nr_1 bits among the N sifted key bits to include all the tagged portion. We make no assumption about the phase errors for the chosen Nr_1 bits. If we count the number of phase errors for the remaining $N' := N(1 - r_1)$ bits, it should be no larger than $N'r_2$ except for a probability $\epsilon_2 := \bar{f}(N'r_2; N', v_{\text{th}}/(L-1))$. The imperfection in the final key is characterized through the failure probability ϵ in identifying the phase error pattern when NH_{PA} bits of error syndrome are given^{10,11,20}. Given $s > 0$, we choose r_1 and r_2 to satisfy $\epsilon_1 = \epsilon_2 = 2^{-s}$. Then, we have $\epsilon \leq 3 \times 2^{-s}$ for $H_{\text{PA}} = r_1 + (1 - r_1)h(r_2) + s/N$. The commonly assumed quality of the key corresponds to about $s = 70$ to 160.

As an example, consider the case with $L = 128$, $\eta = 2 \times 10^{-3}$, and $e_{\text{bit}} = 0.03$. The results below do not change significantly if η is chosen to be smaller, such as $\eta = 10^{-5}$. Asymptotically, the rate is optimized when $\mu = 0.0541$ with $v_{\text{th}} = 17$, leading to $p_1 := e_{\text{src}} N_{\text{em}}/N = 0.047$, $p_2 := v_{\text{th}}/(L-1) = 0.134$, and $H_{\text{PA}} = H_{\text{PA}}^{\text{asy}} := p_1 + (1 - p_1)h(p_2) = 0.588$. For a finite length N , let us first use a crude Gaussian approximation of $\ln[\bar{f}(k; n, p)] \cong -(k - np)^2 / [(2np(1-p))]$. This leads to $r_1 \cong p_1 + \sqrt{(2\ln 2)p_1(s/N)}$ and $r_2 \cong p_2 + \sqrt{(2\ln 2)p_2(1-p_2)(s/N)}$. For $N \gg s$, substituting numerics gives $H_{\text{PA}} \cong H_{\text{PA}}^{\text{asy}}(1 + 1.98\sqrt{s/N})$. A better estimate is given by a rigorous bound $\log_2 \bar{f}(k; n, p) \leq -nD(q||p)$ with $D(q||p) := q \log_2(q/p) + (1-q) \log_2[(1-q)/(1-p)]$. We calculated the finite-size rate after optimizing over μ and v_{th} for each N , and then derived its fraction R to the asymptotic optimal rate of 1.16×10^{-5} . For $s = 100$, we found $R = 3.3\%$ for $N = 10^3$, $R = 53\%$ for $N = 10^4$, $R = 84\%$ for $N = 10^5$, and $R = 95\%$ for $N = 10^6$.

Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat

R. Bintanja¹ & F. M. Selten¹

Precipitation changes projected for the end of the twenty-first century show an increase of more than 50 per cent in the Arctic regions^{1,2}. This marked increase, which is among the highest globally, has previously been attributed primarily to enhanced poleward moisture transport from lower latitudes^{3,4}. Here we use state-of-the-art global climate models⁵ to show that the projected increases in Arctic precipitation over the twenty-first century, which peak in late autumn and winter, are instead due mainly to strongly intensified local surface evaporation (maximum in winter), and only to a lesser degree due to enhanced moisture inflow from lower latitudes (maximum in late summer and autumn). Moreover, we show that the enhanced surface evaporation results mainly from retreating winter sea ice, signalling an amplified Arctic hydrological cycle. This demonstrates that increases in Arctic precipitation are firmly linked to Arctic warming and sea-ice decline. As a result, the Arctic mean precipitation sensitivity (4.5 per cent increase per degree of temperature warming) is much larger than the global value (1.6 to 1.9 per cent per kelvin). The associated seasonally varying increase in Arctic precipitation is likely to increase river discharge^{6–8} and snowfall over ice sheets⁹ (thereby affecting global sea level), and could even affect global climate through freshening of the Arctic Ocean and subsequent modifications of the Atlantic meridional overturning circulation^{10,11}.

Even though the moisture content of the global atmosphere will increase by roughly 7% per degree warming following the Clausius–Clapeyron relation, global average precipitation is projected to increase by only about 2% K^{−1} in response to global warming^{12,13}. This ‘muted’ precipitation response is generally attributed to radiative constraints on the amount of energy available for evaporation¹³. Both observations and climate model simulations consistently show that Arctic precipitation trends far exceed this global average², however. Over the course of the twentieth century, the Arctic experienced an increase in precipitation¹⁰, especially in winter¹⁴, which has largely been attributed to anthropogenic influences¹⁵. Observational uncertainties are considerable¹⁶, however, owing mainly to the paucity of observational data. According to almost all climate models Arctic precipitation will continue to increase and possibly even accelerate in the twenty-first century², with peak increases exceeding 50% (Fig. 1) over the Arctic Ocean.

The potential consequences of such hefty moisture and precipitation increases are widespread: (1) more snow will fall onto adjacent continents¹⁷ and ice sheets⁹, such as the northeastern part of the Greenland ice sheet, which will modulate ice-sheet growth and sea-level change; (2) the ocean surface density will drop owing to a net freshening of the Arctic Ocean (leading, incidentally, to increased stratification that may inhibit sea-ice melt¹⁸), affecting the Atlantic meridional overturning circulation (AMOC) and thereby global climate^{10,11}; (3) river discharge into the Arctic Ocean will become larger, reinforcing the freshening of the Arctic Ocean^{1,6,7,19}; (4) there will be stronger water vapour and cloud feedbacks, reinforcing polar warming^{20–22}; (5) more snowfall on sea ice will potentially affect the ice–ocean albedo feedback¹⁴ and reduce sea-ice growth by increasing insulation; and (6) there are biological and ecological implications, for instance, the greater availability of surface moisture for flora and fauna²³ as Arctic warming causes more precipitation to fall as rain.

The origins of, and mechanisms behind, projected Arctic moistening are still unclear, however. Is it caused by an intensified local Arctic hydrological cycle¹⁹, or does it result from a global hydrological response to altered evaporation rates and moisture fluxes in lower latitudes⁸? This is an important issue, for several reasons: (1) if the local response dominates, Arctic precipitation may be strongly linked to Arctic warming and sea-ice retreat; (2) in that case the effect on Arctic Ocean freshening will probably be limited because evaporation and precipitation effects on surface salinity will then largely cancel out (only a remote origin will lead to overall net freshening); and (3) the local and remote contributions probably exhibit a different seasonal imprint⁴, which is of relevance for all earth system components that depend on the seasonality of precipitation (such as ecology, ice sheets, hydrology)²².

Most studies addressing Arctic precipitation have attributed the projected increase mainly to enhanced poleward moisture transport and convergence^{3,4,8,23}. Global atmospheric warming leads to an increase in atmospheric moisture according to the Clausius–Clapeyron relation. This strengthens meridional moisture gradients, indicating that the moisture transport will increase even if atmospheric dynamics remain invariant⁸. However, climate warming may additionally lead to a poleward shift of the subpolar westerly jet stream³ and to subsequent changes in storm-track location and strength⁹, which affect the magnitude, longitudinal distribution²⁴ and possibly the seasonality of the poleward atmospheric moisture transport, and thereby the contribution of remote source regions. Analyses of water isotopes as a proxy for the origin of Arctic moisture reveal that Arctic precipitation contains a mixture of local and remote origins whose ratio seems to fluctuate seasonally²⁵. The potential contribution of increasing local evaporation (which peaks in autumn/winter²⁶) to Arctic precipitation trends has been addressed generically^{1,10}, but as yet a systematic breakdown of local versus remote contributions is lacking.

Here we use output from 37 state-of-the-art global climate models within the framework of CMIP5 (Coupled Model Intercomparison Project, phase 5)⁵ to systematically quantify projected Arctic precipitation trends, as well as trends in surface evaporation and poleward moisture transport, to assess their relative contributions. We use standardized simulations for the period 2006–2100 based on intermediate and strong forcing scenarios (see Methods Summary). The multimodel-mean twenty-first-century response shows marked relative increases in precipitation and surface evaporation that are largely confined to the Arctic Ocean, with peak values of up to 65% (precipitation) and 300% (evaporation) for the strong forcing scenario (Fig. 1). Moreover, the spatial coherence between peak regions in evaporation and precipitation¹⁶ (for example, in the northern Barents and Kara seas) suggests that surface evaporation changes, most probably reinforced by sea-ice retreat (Fig. 1c), contribute considerably to the amplification of future Arctic precipitation.

The simulated intensification of Arctic (70°–90°N) precipitation varies by a factor of 3 to 4 between models (Fig. 2a) (see Methods). Any increase in Arctic precipitation has only two possible sources: enhanced local surface evaporation (local origin) or intensified moisture transport across 70°N (remote origin). The relative contribution of these sources varies considerably between models (Fig. 2a), but on average surface evaporation

¹Royal Netherlands Meteorological Institute (KNMI), Utrechtseweg 297, 3731 GA, De Bilt, The Netherlands.

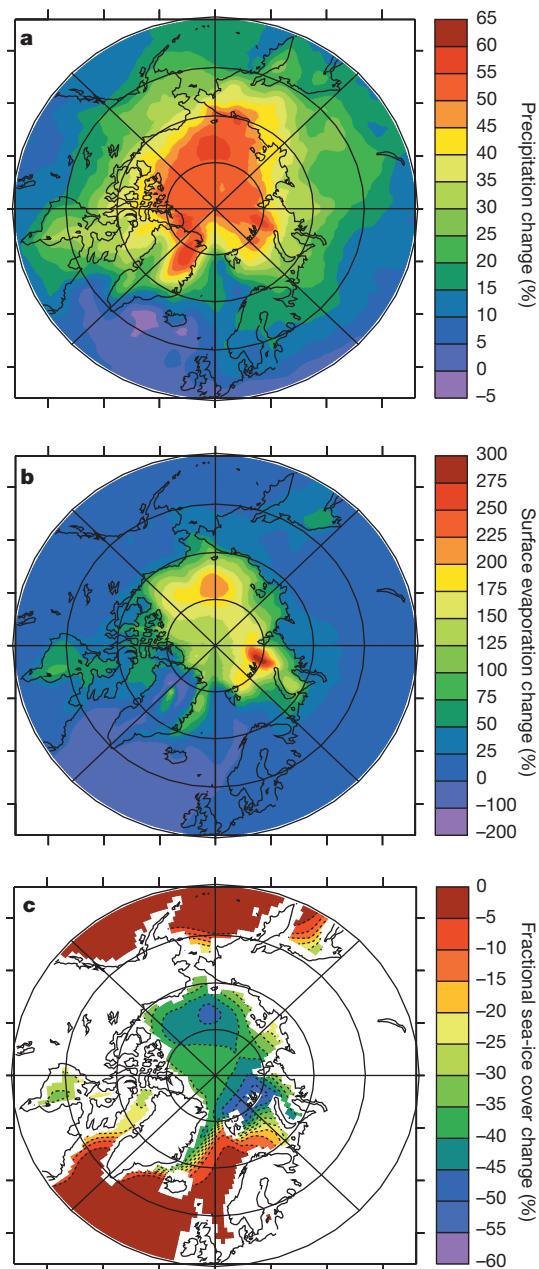


Figure 1 | Geographical distribution of simulated annual mean twenty-first-century precipitation, surface evaporation, and sea-ice cover changes in the Arctic region. Changes (based on the difference between the means over 2091–2100 and 2006–2015) in precipitation (a), surface evaporation (b) and sea-ice extent (c), for the strong forcing scenario (RCP8.5). Precipitation and surface evaporation are presented as relative changes, whereas sea-ice cover is depicted as the absolute change in fractional coverage. Data represent the average of the CMIP5 climate models (see Methods). Simulated geographic patterns are very similar for the intermediate forcing scenario (RCP4.5), with maximum values of 27% relative change for precipitation, 131% relative change for surface evaporation and 28% fractional change for sea-ice cover.

changes provide the major part of the Arctic precipitation increase ($61.9 \pm 20.5\%$ and $56.4 \pm 13.2\%$ of the total for RCP4.5 and RCP8.5, respectively), even though moisture transport contributes as much as two-thirds to Arctic precipitation in the current climate (Methods). Although simulations show that moisture transport towards the Arctic will increase over the twenty-first century, the relative importance of moisture transport to Arctic precipitation will decline with climate warming. Precipitation minus evaporation will become larger, leading to net Arctic Ocean freshening (Methods), but much less than previously conjectured⁴.

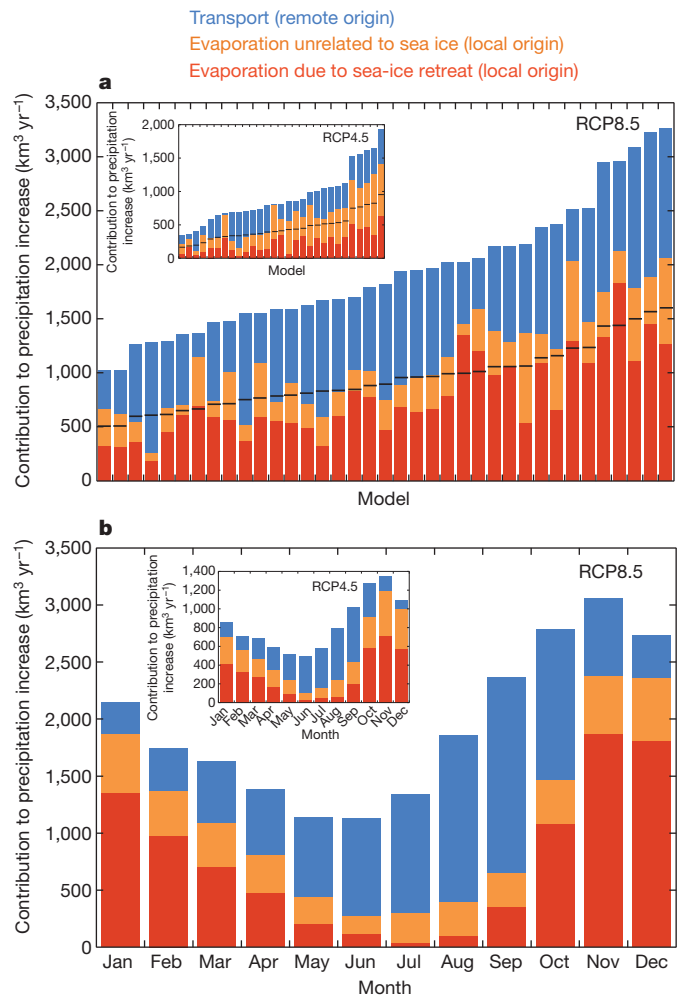


Figure 2 | Simulated annual and monthly twenty-first-century changes in Arctic mean precipitation, poleward moisture transport across 70° N (remote origin) and surface evaporation components (local origin). a, Annual mean changes, where each bar denotes one CMIP5 model sorted according to the magnitude of the simulated precipitation change (horizontal black lines represent 50% of each column). b, Seasonal changes, where each bar represents the monthly and multimodel mean. Changes are given for the strong (RCP8.5) and intermediate (insets, RCP4.5) forcing scenarios. The total Arctic surface evaporation (the sum of the red and orange bars) is separated into its ice-retreat (red) and non-ice-retreat (orange) components. The attribution of sea-ice retreat to the total surface evaporation was evaluated by calculating monthly evaporative flux differences over only the ice-retreat regions, which were then summed over the respective 10-year periods (see Methods Summary).

Sea ice is an important contributor to the enhanced surface evaporation rates²², especially in late autumn and winter (Fig. 2b). When sea ice retreats, the surface consists of open water at the freezing point instead of ice at temperatures far below zero, leading to enormous increases in surface evaporation in the ice-retreat regions^{26,27} (Fig. 1b). As a result, a significant proportion of the total Arctic surface evaporative change can be attributed to sea-ice retreat ($40.9 \pm 11.5\%$ and $68.7 \pm 11.7\%$ for the intermediate and strong forcing scenarios, respectively), especially during the cold winter months (Fig. 2b). Evidently, sea-ice retreat dominates the total evaporative response in the strong forcing scenario, which can be attributed to the projected strong retreat of sea ice ($-48.7 \pm 17.5\%$ relative change in sea-ice area) compared to the intermediate scenario ($-21.1 \pm 12.9\%$). Changes in moisture transport peak in summer and autumn (Fig. 2b), when the increases in meridional temperature and moisture gradients are at their maximum (Arctic temperature changes are minimal in summer). Arctic precipitation increases peak in late autumn

and winter^{1,2} at more than twice the rate of late spring and summer in both scenarios, with strongly seasonally varying contributions from local and remote sources that seem to be consistent with water isotope analyses²⁵.

Sea-ice retreat is obviously related to Arctic warming, which peaks in winter, through the ice–ocean albedo feedback and the seasonal ocean heat storage/release, as well as through other ice-related climate feedbacks^{28,29}. Moreover, increased atmospheric moisture contributes to Arctic warming through the water vapour and cloud feedbacks²¹, and as such constitutes a positive feedback that produces more evaporation through ice retreat²². Because models with strong Arctic warming generally exhibit considerable sea-ice retreat²⁹ (Fig. 3), there is a solid intermodel relation between Arctic warming and changes in surface evaporation (both total and sea-ice-related), in both forcing scenarios (Fig. 3a and b). In contrast, we find only a minor and non-significant intermodel increase in meridional moisture transport changes as a function of Arctic warming (Fig. 3c). This is probably due to counteracting effects such as a smaller meridional temperature gradient associated with Arctic tropospheric warming and a consequent reduction in synoptic activity⁹. Moreover, seasonality may affect this relation because Arctic warming peaks in winter whereas transport changes are largest in summer and early autumn (Fig. 2b). In any case, the dominance of ice retreat in Arctic evaporation increases leads to total precipitation changes being strongly linked to Arctic warming and its associated sea-ice retreat (Fig. 3d), which all peak in late autumn/winter (ref. 29 and Fig. 2b). Models that best simulate the observed sea-ice retreat² exhibit above-average twenty-first-century changes in precipitation and evaporation (Fig. 3), suggesting that the best estimate for these changes would be even higher than the multimodel average. No intermodel relation is found between present-day moisture fluxes and their future changes. Hence, simulated changes in Arctic moisture budget can primarily be attributed to climate feedbacks related to Arctic warming and sea-ice retreat, and only to a lesser degree to the mechanisms behind the intermodel present-day spread in those components.

Enhanced surface evaporation induced by sea-ice retreat thus plays an important part in projected increases in Arctic precipitation. The intensifying local Arctic hydrological cycle significantly amplifies the sensitivity of Arctic mean precipitation ($\Delta P/(P\Delta T) = 4.5\% \text{ K}^{-1}$) over its global mean value (Fig. 4). The primary reason for this amplified precipitation sensitivity is that the evaporation sensitivity in the Arctic is much larger than the global mean, with sea-ice retreat more than doubling its value (Fig. 4). (The non-sea-ice-related evaporation sensitivity $\Delta E/(E\Delta T)$ is already larger than its global mean owing mainly to the reduced vertical stability of the lower troposphere, especially in the non-summer months²⁶). Simulated atmospheric moisture transport sensitivities are relatively small ($\sim 3\% \text{ K}^{-1}$) compared to the Clausius–Clapeyron relation ($6\text{--}7\% \text{ K}^{-1}$), which might indicate that subdued atmospheric dynamics and diminished meridional temperature gradients effectively reduce the moisture transport sensitivity.

Projected twenty-first-century increases in Arctic precipitation rates of over 50% are among the highest globally, a consistent feature among state-of-the-art global climate models. Precipitation changes peak in late autumn and winter, with the local and remote contributions exhibiting a pronounced seasonal cycle. Arctic warming and sea-ice retreat intensify the local hydrological cycle within the Arctic region through increasing the amount of open water and amplifying surface evaporation (Fig. 3b), especially in winter, causing strongly increased precipitation rates, in contrast to earlier findings⁴. The Arctic mean precipitation sensitivity (relative change per degree of warming) of about $4.5\% \text{ K}^{-1}$ is much larger than the global mean sensitivity of $1.6\text{--}1.9\% \text{ K}^{-1}$, owing mainly to Arctic warming and sea-ice retreat. Evidently, the local hydrological cycle intensifies at the expense of the large-scale hydrological cycle in increasing Arctic precipitation rates.

The associated breakdown between local ($\sim 60\%$, peaking in late autumn and winter) and remote ($\sim 40\%$, peaking in summer and early autumn) origins of Arctic precipitation changes is important for understanding future Arctic Ocean surface salinity trends (Methods); this is because only precipitation from remote origins leads to net freshening

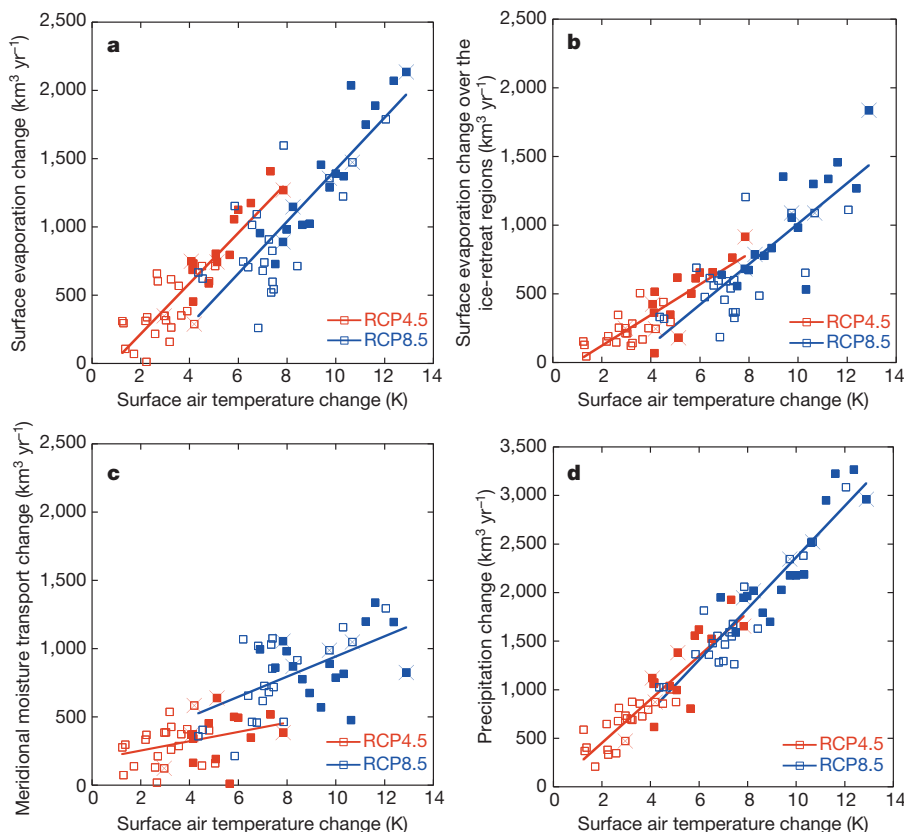


Figure 3 | Dependence of simulated Arctic mean twenty-first-century changes in surface evaporation, poleward moisture transport and precipitation on Arctic mean surface air temperature change. Relations are given for total surface evaporation (a), surface evaporation in the ice-retreat regions (b), poleward moisture transport across 70° N (c) and precipitation (d). Red and blue squares denote the intermediate and strong forcing scenarios, respectively, with each square representing one CMIP5 model. Open (or full) squares represent models with sea-ice retreat smaller (or larger) than the multi-model average. The five models that best simulate the observed sea-ice trends over the past decades² are indicated by crosses. The straight lines are the respective linear regressions.

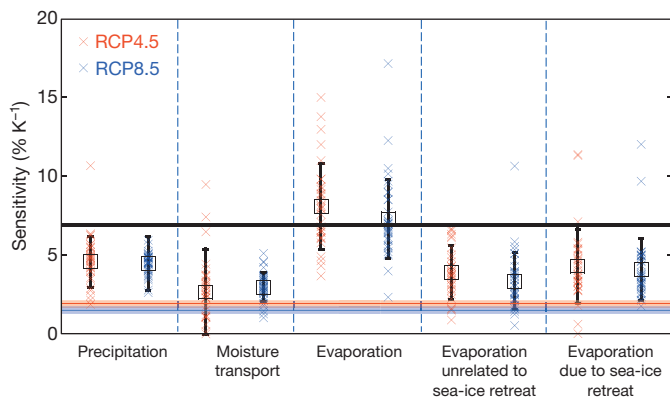


Figure 4 | Simulated twenty-first-century Arctic mean precipitation, evaporation and moisture transport sensitivities. The sensitivity in a variable X , where X is either precipitation (P), evaporation (E) or moisture transport (F), is defined as $\Delta X/(X\Delta T)$, in which T is the Arctic mean temperature (note that for the evaporation components the denominator uses total evaporation). Crosses denote individual models. Black open squares indicate model-mean values, and vertical black error bars represent the associated intermodel standard deviations. The red and blue horizontal lines show the model-mean and global-mean precipitation sensitivities (which equals the global-mean evaporation sensitivity), with the surrounding band denoting the associated intermodel standard deviations. Results are given for the intermediate (red, $n = 33$) and strong (blue, $n = 37$) forcing scenarios. The black horizontal line represents the Clausius–Clapeyron relation.

of the Arctic Ocean (in conjunction with increased river discharge^{1,6,7,19}). This, in turn, will affect the Atlantic ocean circulation (Methods) and thereby global climate, and counteract sea-ice melt¹⁸. The seasonal imprint of precipitation increases affects snow accumulation on sea ice (affecting its seasonal growth and decay) as well as on ice sheets such as north-eastern Greenland (Fig. 1a) and adjacent continental areas⁹ (because the peak increase occurs during winter most additional precipitation will fall as snow), with possible modulations of global sea level and seasonal snow albedo feedbacks¹⁴.

METHODS SUMMARY

In all analyses we used the Coupled Model Intercomparison Project, phase 5 (CMIP5) state-of-the-art global climate models⁵, which were applied in a series of standardized forcing scenarios for the period 2006–2100. Here we use the intermediate (RCP4.5) and strong (RCP8.5) forcing scenarios, for which the combined greenhouse, aerosol and other radiative forcings in the year 2100 totals 4.5 W m^{-2} and 8.5 W m^{-2} , respectively⁵. We use all models for which data coverage was complete and without obvious errors (other than that no selection of models was made); one ensemble member per model (the first) was used. Twenty-first-century trends in the Arctic moisture budget terms are defined here as the difference between the means over the periods 2091–2100 and 2006–2015, with all uncertainties throughout the paper being defined as the standard deviations of the model ensemble.

In this study we define the Arctic as the region 70° – 90° N because most of the precipitation and sea-ice changes occur in the Arctic Ocean north of 70° N (Fig. 1). The moisture budget of the Arctic atmosphere can be expressed as $\partial Q/\partial t = F - (P - E)$, where Q is the total moisture content of the atmosphere (the sum of all atmospheric water vapour, water and ice), P is the total Arctic precipitation (downward), E is the total Arctic surface evaporation (upward) and F is the poleward atmospheric moisture transport across 70° N. The Arctic atmosphere moisture reservoir is quite small; the mean residence time of atmospheric moisture is less than one week, meaning that $\partial Q/\partial t \ll F, P, E$. We thus evaluate the poleward transport using the balance: $F = P - E$.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 November 2013; accepted 11 March 2014.

Published online 7 May 2014.

1. Kattsov, V. M. *et al.* Simulation and projection of Arctic freshwater budget components by the IPCC AR4 global climate models. *J. Hydrometeorol.* **8**, 571–589 (2007).

2. Collins, M. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 1029–1136 (Cambridge Univ. Press, 2013).
3. Kug, J. S. *et al.* Role of synoptic eddy feedback on polar climate responses to the anthropogenic forcing. *Geophys. Res. Lett.* **37**, L14704 (2010).
4. Bengtsson, L. *et al.* The changing atmospheric water cycle in polar regions in a warmer climate. *Tellus A* **63**, 907–920 (2011).
5. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
6. Peterson, B. J. *et al.* Increasing river discharge to the Arctic Ocean. *Science* **298**, 2171–2173 (2002).
7. Holland, M. M., Finnis, J. & Serreze, M. C. Simulated Arctic Ocean freshwater budgets in the twentieth and twenty-first centuries. *J. Clim.* **19**, 6221–6242 (2006).
8. Zhang, X. *et al.* Enhanced poleward moisture transport and amplified northern high-latitude wetting trend. *Nature Clim. Change* **3**, 47–51 (2013).
9. Singarayer, J. S., Bamber, J. L. & Valdes, P. J. Twenty-first-century climate impacts from a declining Arctic sea ice cover. *J. Clim.* **19**, 1109–1125 (2006).
10. Kattsov, V. M. & Walsh, J. E. Twentieth-century trends of Arctic precipitation from observational data and a climate model simulation. *J. Clim.* **13**, 1362–1370 (2000).
11. Davies, F. J., Renssen, H. & Goosse, H. The Arctic freshwater cycle during a naturally and an anthropogenically induced warm climate. *Clim. Dyn.* **42**, 2099–2112 (2014).
12. Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).
13. Stephens, G. L. & Ellis, T. D. Controls of global-mean precipitation increases in global warming GCM experiments. *J. Clim.* **21**, 6141–6155 (2008).
14. Screen, J. A. & Simmonds, I. Declining summer snowfall in the Arctic: causes, impacts and feedbacks. *Clim. Dyn.* **38**, 2243–2256 (2012).
15. Min, S. K., Zhang, X. & Zwiers, F. Human-induced Arctic moistening. *Science* **320**, 518–520 (2008).
16. Walsh, J. E. *et al.* Arctic precipitation and evaporation: model results and observational estimates. *J. Clim.* **11**, 72–87 (1998).
17. Liu, J. *et al.* Impact of declining Arctic sea ice on winter snowfall. *Proc. Natl Acad. Sci. USA* **109**, 4074–4079 (2012).
18. Bintanja, R. *et al.* Important role for ocean warming and increased ice-shelf melt in Antarctic sea-ice expansion. *Nature Geosci.* **6**, 376–379 (2013).
19. Holland, M. M. *et al.* Projected changes in Arctic Ocean freshwater budgets. *J. Geophys. Res.* **112**, G04S55 (2007).
20. Cai, M. Dynamical amplification of polar warming. *Geophys. Res. Lett.* **32**, L22710 (2005).
21. Alexeev, V. A., Langen, P. L. & Bates, J. R. Polar amplification of surface warming on an aquaplanet in ‘ghost forcing’ experiments without sea ice feedbacks. *Clim. Dyn.* **24**, 655–666 (2005).
22. Francis, J. A. *et al.* An arctic hydrologic system in transition: feedbacks and impacts on terrestrial, marine, and human life. *J. Geophys. Res.* **114**, G04019 (2009).
23. ACIA *Arctic Climate Impact Assessment* (Cambridge Univ. Press, 2005).
24. Walsh, J. E. in *The Freshwater Budget of the Arctic Ocean* 21–43 (eds Lewis, E. L. *et al.*) (Kluwer, 2000).
25. Kurita, N. Origin of Arctic water vapor during the ice-growth season. *Geophys. Res. Lett.* **38**, L02709 (2011).
26. Bintanja, R., Graversen, R. G. & Hazeleger, W. Arctic winter warming amplified by the thermal inversion and consequent low infrared cooling to space. *Nature Geosci.* **4**, 758–761 (2011).
27. Barry, R. G. & Serreze, M. C. in *The Freshwater Budget of the Arctic Ocean* 45–56 (eds Lewis, E. L. *et al.*) (Kluwer, 2000).
28. Bintanja, R., van der Linden, E. C. & Hazeleger, W. Boundary layer stability and Arctic climate change: a feedback study using EC-Earth. *Clim. Dyn.* **39**, 2659–2673 (2012).
29. Bintanja, R. & van der Linden, E. C. The changing seasonal climate in the Arctic. *Sci. Rep.* **3**, 1556 (2013).

Acknowledgements We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank all climate-modelling groups for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We are grateful to the EC-Earth consortium for their contribution to the development of the Earth System Model EC-Earth. We thank C. A. Katsman and R. G. Graversen for their comments on the manuscript, and to G. J. van Oldenborgh for information on intermodel versus intramodel climate variability.

Author Contributions R.B. developed the ideas that led to this paper. R.B. and F.M.S. analysed the climate model simulations. R.B. wrote the main paper, with input from F.M.S. Both authors discussed the results and implications and commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (bintanja@knmi.nl or bintanja@gmail.com).

METHODS

Schematic of main results. The main conclusions of this paper have been summarized in Extended Data Fig. 1. It schematically depicts the twenty-first-century changes in the simulated hydrological cycle of the Arctic region, as well as the most important climatological consequences of these changes.

Intermodel differences in moisture budget components. Here we present current and future CMIP5 (ref. 5) model-mean values of total Arctic precipitation, surface evaporation and the poleward moisture transport across 70°N for the intermediate forcing scenario RCP4.5 (Extended Data Table 1) and the strong forcing scenario RCP8.5 (Extended Data Table 2). We note that while the changes in local surface evaporation dominate the changes in Arctic precipitation, moisture transport is the largest term at present and this remains the case in the future, both for RCP4.5 and RCP8.5. To be able to compare the simulated present-day moisture budget of the Arctic with those evaluated from reanalyses data⁴, we also calculated the moisture budget components for the Arctic defined as 60° – 90°N . In the CMIP5 model ensemble (RCP4.5), present-day annual and Arctic mean (60° – 90°N) moisture fluxes are: $P = 17,368 \pm 1,911 \text{ km}^3 \text{ yr}^{-1}$, $E = 7,961 \pm 1,133 \text{ km}^3 \text{ yr}^{-1}$ and $F = 9,407 \pm 1,989 \text{ km}^3 \text{ yr}^{-1}$. These are very close to present-day values evaluated from reanalyses data⁴: $P = 17,408 \text{ km}^3 \text{ yr}^{-1}$, $E = 8,073 \text{ km}^3 \text{ yr}^{-1}$ and $F = 8,830 \text{ km}^3 \text{ yr}^{-1}$.

Extended Data Fig. 2 shows distributions of modelled precipitation, evaporation and moisture transport across 60°N . Evidently, even though the multimodel averages are quite close to the reanalyses-derived values, there is substantial spread among models concerning the Arctic moisture budget components. The figures also depict the distributions for the future scenarios (RCP4.5 and RCP8.5), which exhibit an even larger spread. Interestingly, though, the intermodel spread in the present-day moisture budget components seems to have little bearing on the simulated twenty-first-century changes in these components (and the spread therein) because we find no appreciable intermodel correlations between the present-day values and their changes. Instead, the changes in moisture budget components are governed by Arctic warming and sea-ice retreat (see Fig. 3). Hence, we conclude that the simulated changes in Arctic moisture budget can primarily be attributed to climate feedbacks related to Arctic warming and sea-ice retreat, and only to a lesser degree to the mechanisms behind the intermodel present-day spread in those components.

Effects of Arctic precipitation changes on sea surface salinity. A potentially far-reaching consequence of increased precipitation in the Arctic region is its effect on Arctic Ocean salinity, primarily because of its potential relation to (global) ocean currents (see next section). Extended Data Fig. 3 depicts simulated twenty-first-century changes in average (70° – 90°N) Arctic Ocean sea surface salinity $S_{\text{sea-surface}}$ in the CMIP5 models for the intermediate (RCP4.5) and strong (RCP8.5) forcing scenarios, respectively. With a few exceptions in the intermediate RCP4.5 scenario, all models exhibit an overall decrease in $S_{\text{sea-surface}}$, consistent with the simulated increases in Arctic precipitation. However, a number of other climate mechanisms also potentially affect Arctic Ocean $S_{\text{sea-surface}}$: local evaporation, increased river discharge and runoff⁶, sea-ice melt and altered ocean advection and mixing¹⁸. Hence, Arctic Ocean freshening cannot be straightforwardly attributed to increases in precipitation.

To quantify the effect of Arctic precipitation increase on Arctic sea surface salinity we carried out additional climate model simulations using the global coupled climate model EC-Earth³⁰, which is part of the CMIP5 model ensemble. For this purpose we use exactly the same model settings as in the CMIP5 simulations that were carried out with EC-Earth. Two 44-year simulations were done for perpetual year 2006 forcing conditions: a control simulation, and a simulation in which the precipitation over the Arctic Ocean north of 70°N is artificially increased by 50%. Even though the climate will drift somewhat because of the applied artificial increase in precipitation (for instance, sea ice is found to expand), the difference between these two simulations will at least provide a rough estimate of how much increased precipitation contributes to changes in Arctic Ocean $S_{\text{sea-surface}}$. Extended Data Fig. 4 shows the simulated difference in the geographical distribution of S . Although there is a general decrease in $S_{\text{sea-surface}}$ over most of the Arctic Ocean (as well as in the North Atlantic), there is a region where $S_{\text{sea-surface}}$ exhibits an increase. The simulation shows an overall increase in sea ice (not shown), which may be due to the

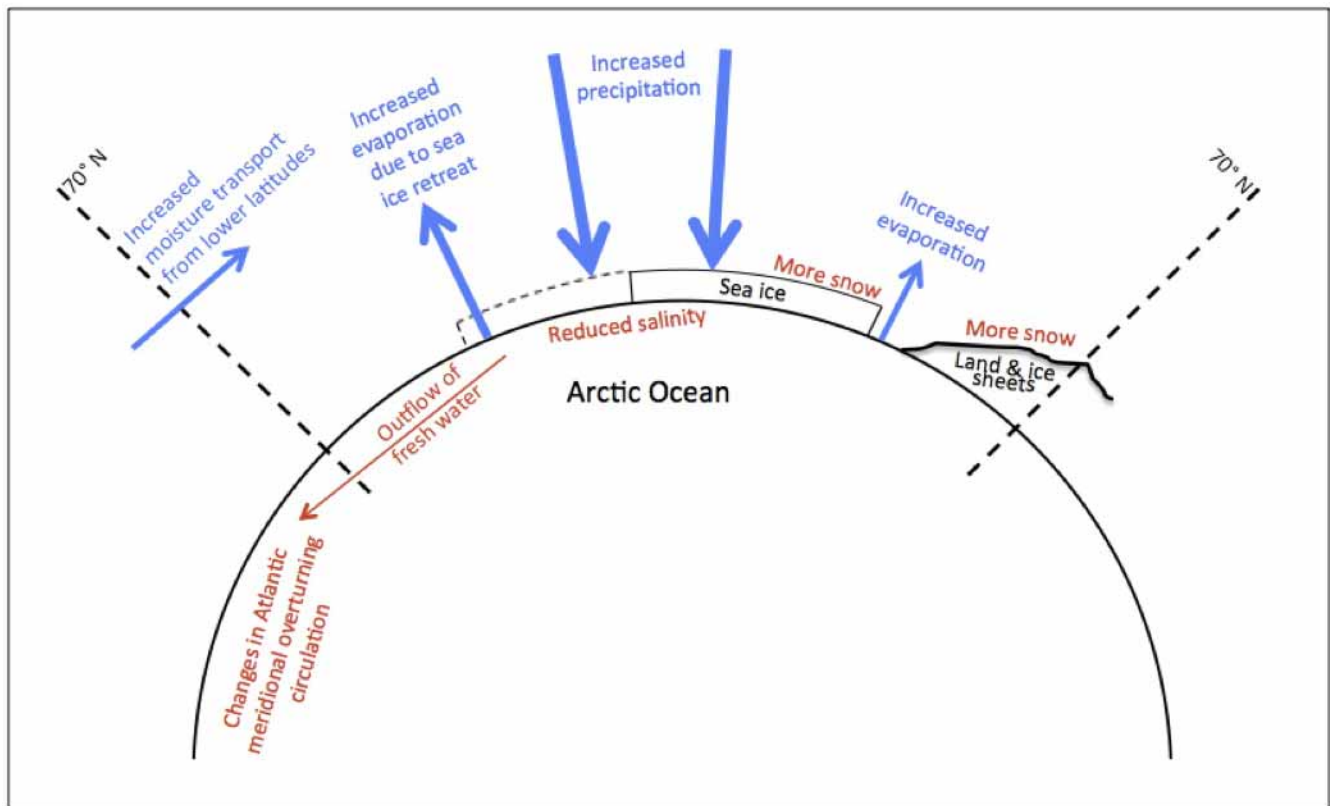
positive feedback of overall freshening on sea ice formation, or to internal climate variability. In regions with strong sea-ice expansion the associated brine rejection is apparently able to counteract the overall $S_{\text{sea-surface}}$ decrease caused directly by increased precipitation (simulations with larger artificial precipitation increases exhibit a decrease in $S_{\text{sea-surface}}$ throughout the entire Arctic Ocean). With a few simplifying assumptions (for example, that changes in $S_{\text{sea-surface}}$ scale linearly with those in the applied precipitation increase and that only remote moisture leads to net freshening) we estimate the contribution of increases in Arctic precipitation to Arctic Ocean surface freshening at about 25% for both the RCP4.5 and RCP8.5 scenarios. Hence, according to these simulations, about a quarter of the decrease in $S_{\text{sea-surface}}$ can be attributed to increased precipitation in the Arctic.

To further assess the effect of local versus remote moisture sources on Arctic Ocean $S_{\text{sea-surface}}$, we show in Extended Data Fig. 5 the intermodel relation between the ratio of the changes in moisture transport across 70°N and in surface evaporation (that is, the ratio of remote and local contributions), and simulated $S_{\text{sea-surface}}$ changes. The correlation is small (-0.63), presumably because of other mechanisms also affecting Arctic Ocean $S_{\text{sea-surface}}$, as mentioned above. Nevertheless, the results suggest that models with a relatively large increase in moisture transport (remote sources) exhibit relatively strong reductions in $S_{\text{sea-surface}}$, whereas models with strong Arctic evaporation increase (local sources) simulate a relatively small decrease in $S_{\text{sea-surface}}$. Hence, moisture originating from remote sources has the largest effect on mean Arctic Ocean $S_{\text{sea-surface}}$; to understand changes in Arctic Ocean $S_{\text{sea-surface}}$ it is therefore important to know whether the increase in Arctic precipitation is due to moisture transport from remote sources or to local surface evaporation.

Effects of Arctic precipitation changes on the AMOC. The CMIP5 climate runs (RCP4.5, RCP8.5) exhibit a significant reduction in AMOC in almost all climate models², which is generally attributed to reduced surface ocean density in the deep-water formation regions in the North Atlantic (either by an increase in sea surface temperatures or by a decrease in $S_{\text{sea-surface}}$). Increases in Arctic precipitation and surface freshening of the Arctic Ocean and the associated outflow of this relatively fresh and light surface water towards the North Atlantic will potentially contribute to the AMOC decrease⁷. Here we investigate this contribution through a series of 44-year simulations using the global climate model EC-Earth³⁰, in which we artificially altered the precipitation rates over the Arctic Ocean north of 70°N , similar to the simulations described in the previous section. Arctic precipitation rate changes of -50% , $+50\%$, $+100\%$ and $+300\%$ were applied in separate simulations, in addition to a control run without any change. In the previous section we showed that, as expected, increases in Arctic precipitation lead to an average surface freshening in the Arctic Ocean, and also in the North Atlantic Ocean (Extended Data Fig. 4).

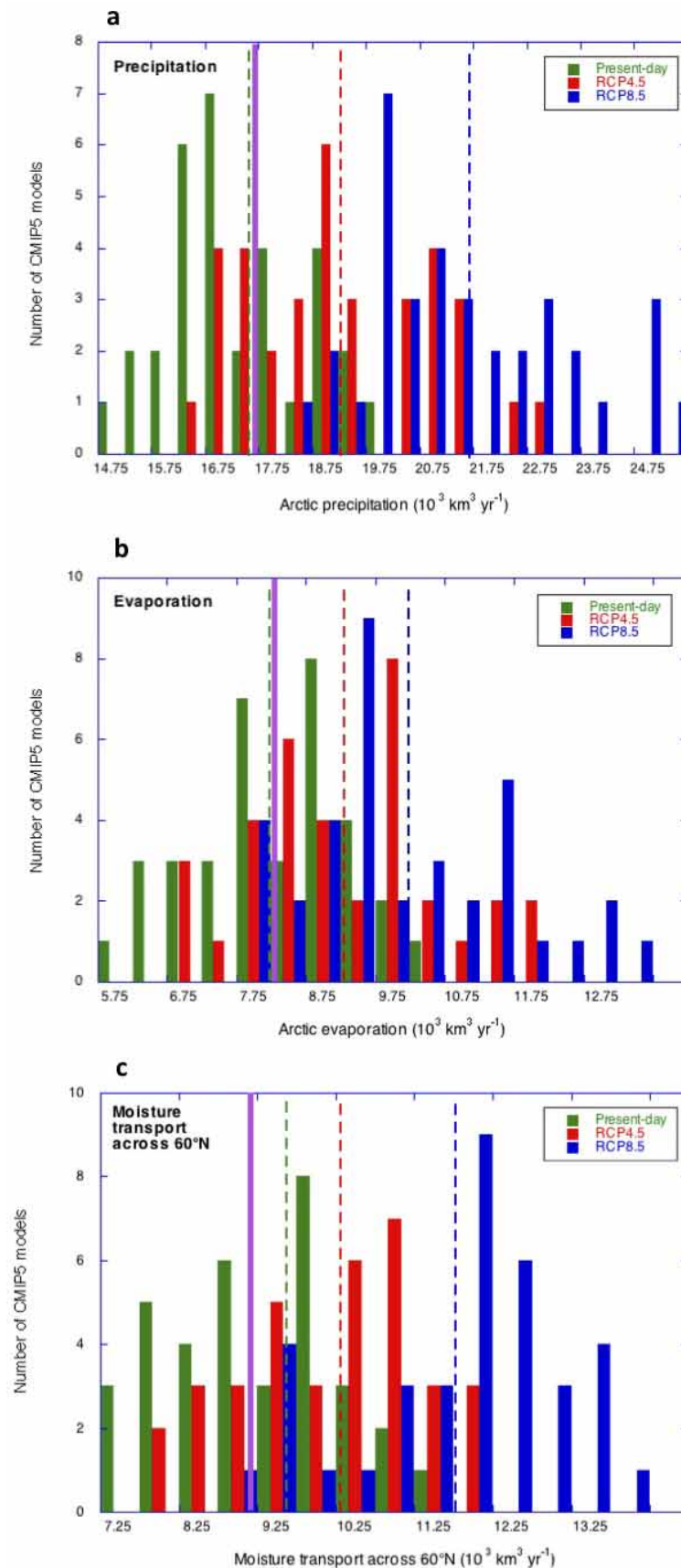
In Extended Data Fig. 6 we show the temporal variation in annual mean AMOC strength at 30°N (definition used in ref. 2) for the various Arctic precipitation perturbation simulations. Evidently, increases in Arctic precipitation suppress the AMOC strength through exporting relatively fresh water to the North Atlantic, causing a reduced ocean surface salinity and density in the deep-convection zones. Averaged over the final 20 years, the control AMOC strength at 30°N equals 12.8 ± 0.4 Sverdrup (Sv), whereas in the $+50\%$ Arctic Ocean precipitation simulation this value is 11.4 ± 0.6 Sv (the uncertainty represents the standard deviation based on interannual variations in AMOC strength). Hence, according to these simulations, a 50% increase in Arctic precipitation leads to a significant decrease in AMOC strength of about 10%. Given that the model-average AMOC decrease over the twenty-first century in the CMIP5 model ensemble for the RCP8.5 scenario is about 40% (ref. 2), and in EC-Earth only about 25%, (owing mainly to its comparatively weak AMOC in the present-day climate) the projected increase in Arctic precipitation and the associated outflow of fresh surface water towards the North Atlantic contributes substantially to the projected decrease in the AMOC over the twenty-first century.

30. Hazeleger, W. *et al.* EC-Earth: a seamless Earth system prediction approach in action. *Bull. Am. Meteorol. Soc.* **91**, 1357–1363 (2010).



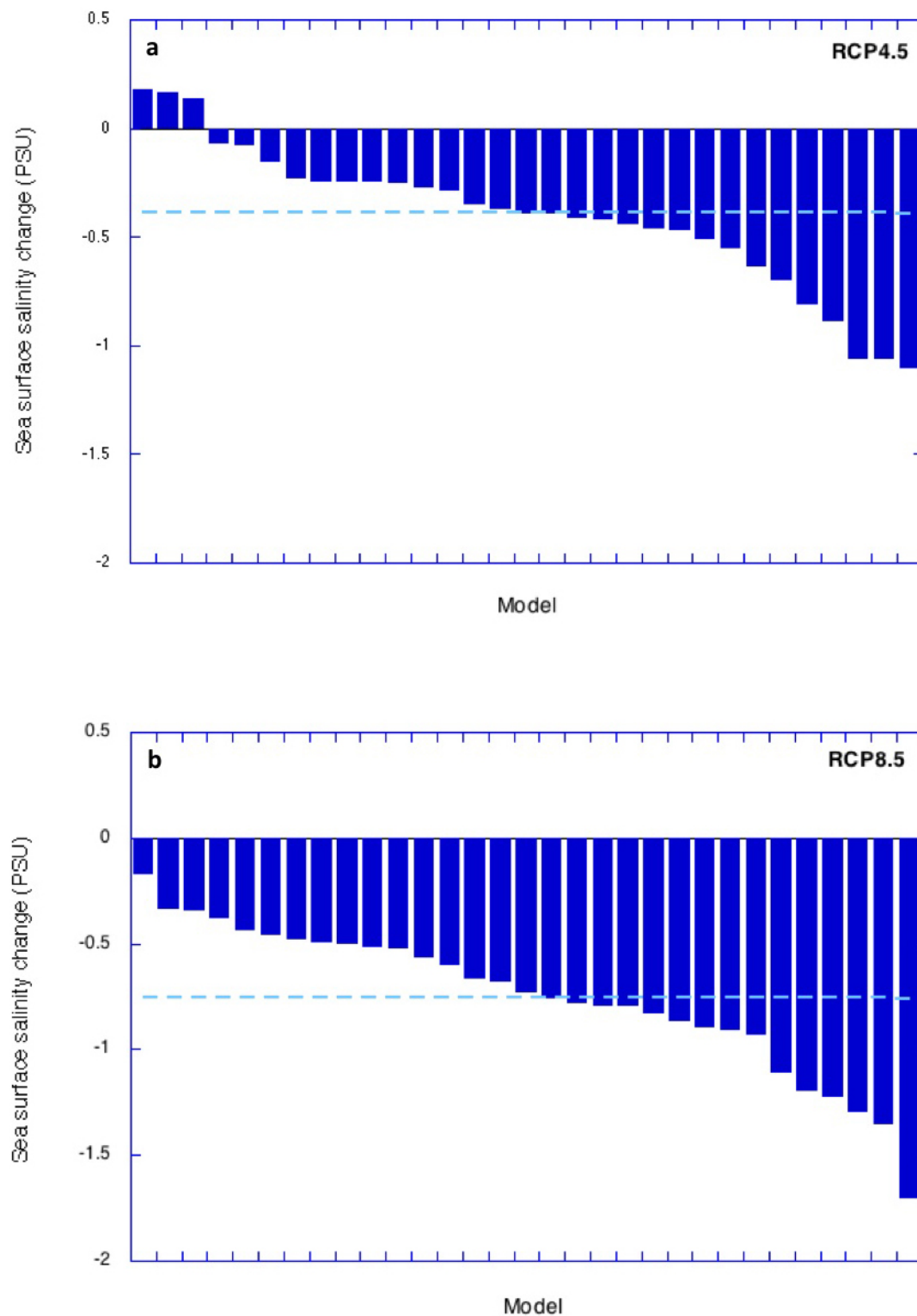
Extended Data Figure 1 | Schematic overview of the main processes discussed in the paper. Blue arrows denote future changes in the Arctic various moisture fluxes, with the size of the arrows loosely depicting the

magnitude of the changes. The red text and arrows represent the associated climatic consequences.



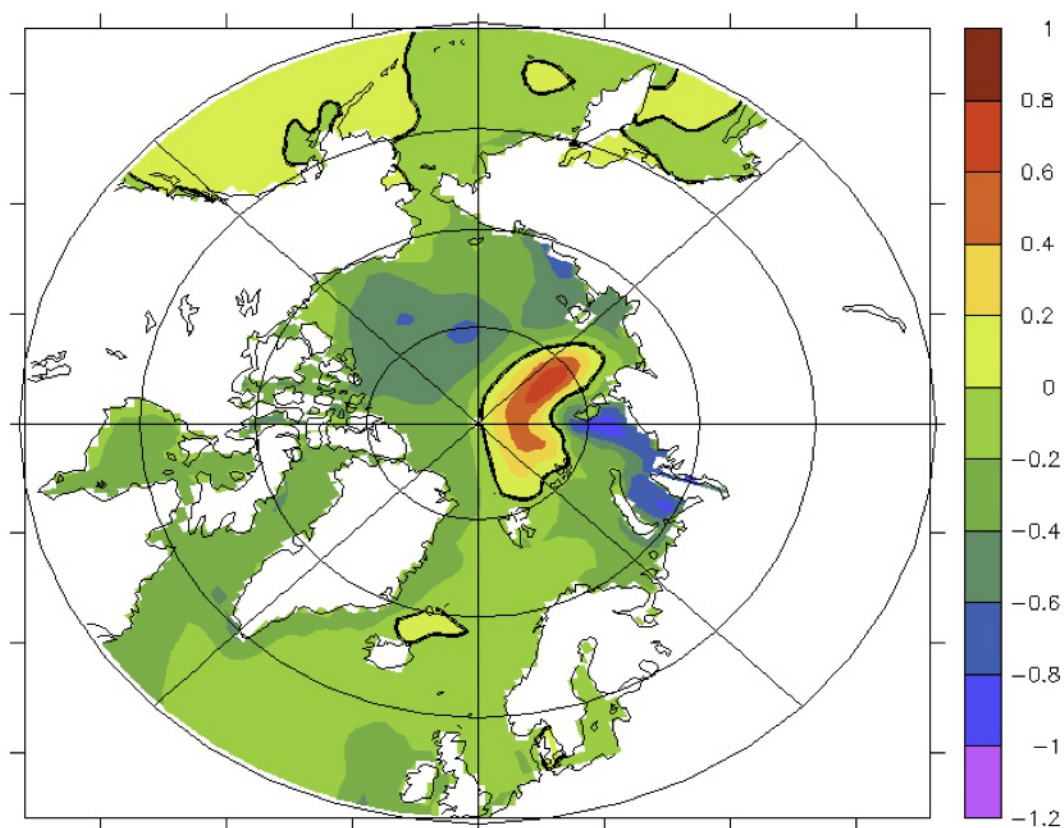
Extended Data Figure 2 | Number of CMIP5 models that simulate specific amounts of precipitation, surface evaporation and moisture transport across 60°N. Distributions of Arctic (60°–90°N) precipitation (a), surface evaporation (b) and poleward moisture transport across 60°N (c) in terms of

the number of CMIP5 models that simulate a specific amount. Distributions are given for the present (green), future RCP4.5 (red) and future RCP8.5 (blue). Dashed lines denote the model-mean values, while the purple line shows the present-day value derived from reanalyses data⁴.

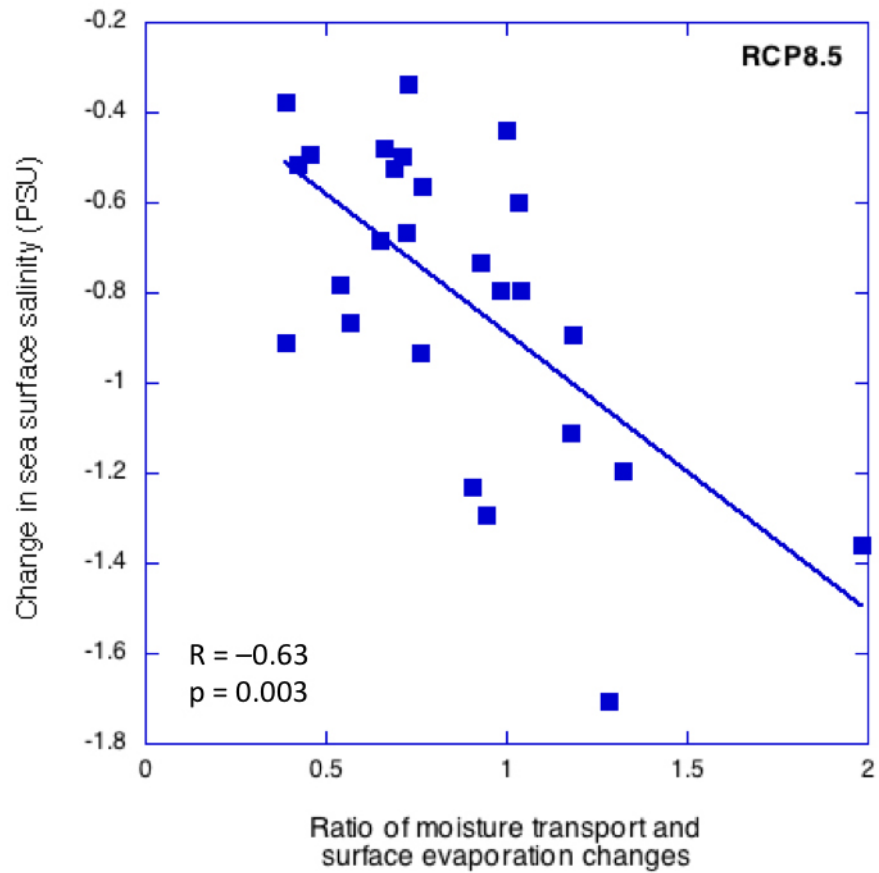


Extended Data Figure 3 | Simulated twenty-first-century changes in sea surface salinity in the Arctic Ocean north of 70° N. Results are shown for the intermediate forcing scenario (RCP4.5) (a) and the strong forcing (RCP8.5)

(b), sorted according to the magnitude of the change. The dashed line shows the multimodel-mean value. The salinity is shown on the colour scale in practical salinity units (PSU).

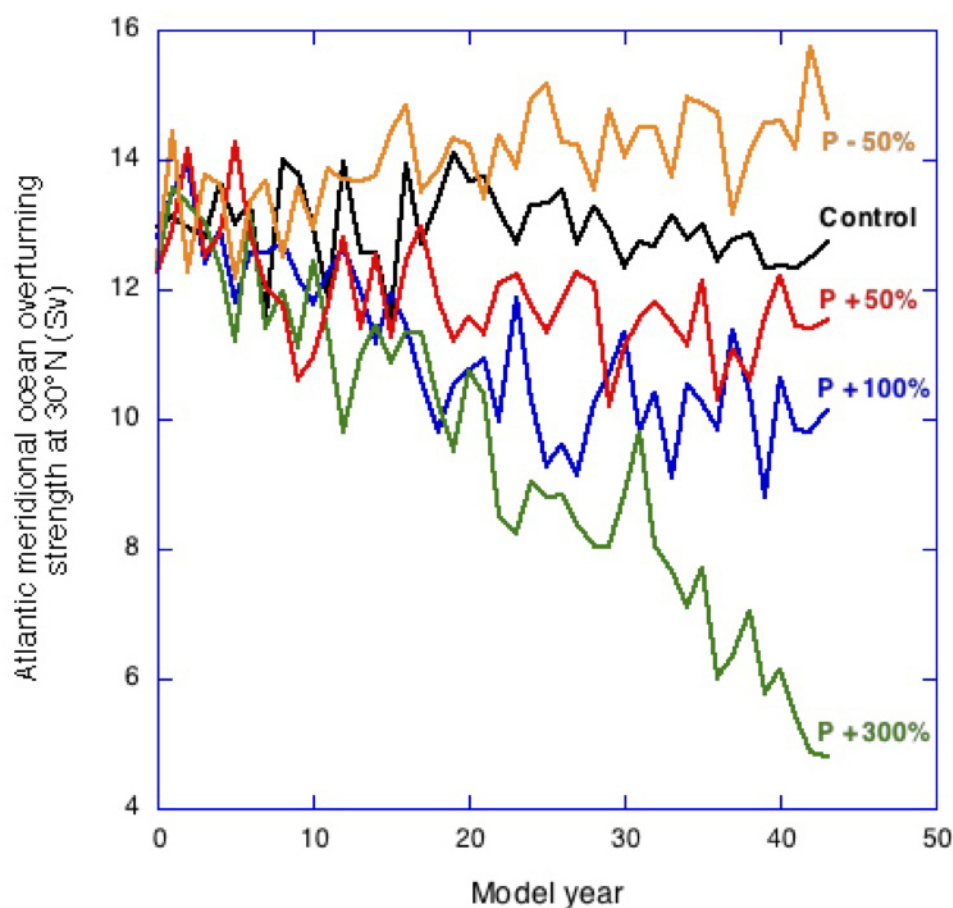


Extended Data Figure 4 | Simulated changes in sea surface salinity resulting from a 50% increase in precipitation over the Arctic Ocean north of 70° N using the global climate model EC-Earth³⁰. The salinity change is shown on the colour scale in PSU.



Extended Data Figure 5 | Simulated change in Arctic mean sea surface salinity against the ratio of changes in moisture transport across 70° N and surface evaporation. The ratio is equivalent to the ratio of remote and local contributions. Each square represents one CMIP5 model. The straight line

shows the best linear fit to the data, where R is the correlation coefficient and p is the probability that uncorrelated data of the same sample size would yield a similar relationship.



Extended Data Figure 6 | Simulated AMOC strength as a function of Arctic precipitation increase. Time series of simulated annual-mean AMOC

strength at 30° N for the various Arctic precipitation (P) change simulations (see main text for details) carried out with the global climate model EC-Earth³⁰.

Extended Data Table 1 | Arctic moisture flux components and their changes for the RCP4.5 scenario

	Present-day	Future	Change (abs.)	Change (%)
Precipitation	5162 ± 784	5990 ± 833	827 ± 478	+16
Evaporation	1857 ± 417	2389 ± 651	532 ± 382	+29
Transport	3305 ± 788	3678 ± 756	296 ± 182	+9

Total Arctic (70°–90° N) moisture flux components for the present (2006–2015) and the future (2091–2100) based on the RCP4.5 forcing scenario, as well as the absolute (abs.) and relative twenty-first-century changes. Values are based on the CMIP5 model ensemble (see Methods Summary), with the uncertainties representing the standard deviations of the multimodel ensemble. Units are $\text{km}^3 \text{yr}^{-1}$.

Extended Data Table 2 | Arctic moisture flux components and their changes for the RCP8.5 scenario

	Present-day	Future	Change (abs.)	Change (%)
Precipitation	5175 ± 710	7118 ± 1001	1943 ± 600	+38
Evaporation	1840 ± 418	3042 ± 718	1118 ± 480	+61
Transport	3335 ± 736	4160 ± 855	825 ± 277	+25

Total Arctic (70°–90° N) moisture flux components for the present (2006–2015) and the future (2091–2100) based on the RCP8.5 forcing scenario, as well as the absolute and relative twenty-first-century changes. Values are based on the CMIP5 model ensemble (see Methods Summary), with the uncertainties representing the standard deviations of the multimodel ensemble. Units are $\text{km}^3 \text{yr}^{-1}$.

Uplift and seismicity driven by groundwater depletion in central California

Colin B. Amos¹, Pascal Audet², William C. Hammond³, Roland Bürgmann^{4,5}, Ingrid A. Johanson⁴ & Geoffrey Blewitt³

Groundwater use in California's San Joaquin Valley exceeds replenishment of the aquifer, leading to substantial diminution of this resource^{1–4} and rapid subsidence of the valley floor⁵. The volume of groundwater lost over the past century and a half also represents a substantial reduction in mass and a large-scale unburdening of the lithosphere, with significant but unexplored potential impacts on crustal deformation and seismicity. Here we use vertical global positioning system measurements to show that a broad zone of rock uplift of up to 1–3 mm per year surrounds the southern San Joaquin Valley. The observed uplift matches well with predicted flexure from a simple elastic model of current rates of water-storage loss, most of which is caused by groundwater depletion³. The height of the adjacent central Coast Ranges and the Sierra Nevada is strongly seasonal and peaks during the dry late summer and autumn, out of phase with uplift of the valley floor during wetter months. Our results suggest that long-term and late-summer flexural uplift of the Coast Ranges reduce the effective normal stress resolved on the San Andreas Fault. This process brings the fault closer to failure, thereby providing a viable mechanism for observed seasonality in microseismicity at Parkfield⁶ and potentially affecting long-term seismicity rates for fault systems adjacent to the valley. We also infer that the observed contemporary uplift of the southern Sierra Nevada previously attributed to tectonic or mantle-derived forces^{7–10} is partly a consequence of human-caused groundwater depletion.

Hydrospheric mass changes exert direct influence over lithospheric deformation. Both seasonal and long-term changes to ice, snow or water loads may induce displacements of the Earth's surface^{11–15} and can create stress perturbations that modulate activity on seismogenic faults^{16–19}. A volume of groundwater approaching approximately 160 km³ in California's Central Valley has been lost through pumping, irrigation and evapotranspiration over the past 150 years or so^{3,4}. Historical and modern records demonstrate that groundwater depletion occurs primarily in the drier, hotter, southern portion of the basin^{1,3} (the San Joaquin Valley), parallel to the central San Andreas Fault and adjacent to the high topography of the southern Sierra Nevada (Fig. 1). Previous studies demonstrating sensitivity to small-scale stress changes across this section of the San Andreas point to seasonal hydrologic⁶ or temperature²⁰ variations to explain observed changes in seismicity rates⁶ or strain in shallow boreholes²⁰.

Modest contemporary uplift rates of the southern Sierra Nevada observed using space geodesy^{7–9} have been attributed to various tectonic and geomorphic drivers, such as buoyant response to mantle delamination¹⁰, Basin and Range extensional faulting²¹, and erosional mass transfer from Pleistocene glaciation²². Inferred stress change from epeirogenic uplift in the southern Sierra Nevada is also potentially linked to the transition from locked to creeping sections of the San Andreas Fault²³. Here we seek to explore potential human impacts on contemporary deformation across this region through global positioning system (GPS) constraints on lithospheric flexure induced by anthropogenic groundwater level changes.

Continuous GPS networks spanning the southwestern USA provide a high-resolution framework for analysing crustal motion in three dimensions (Fig. 1). Longer available records and recently improved processing techniques enable determination of reliable GPS vertical velocities with a precision²⁴ of less than a millimetre per year, tied to the Earth system's centre of mass to within 0.5 mm yr^{–1} (refs 24 and 25). GPS station data were processed to create individual time series of vertical position and were then fitted with an empirical model including epoch position, velocity, and the amplitude and phase of annual and semiannual harmonic components (to model seasonal effects). This analysis includes only stations with at least two-and-a-half years of data and vertical rate uncertainties of ≤ 1.0 mm yr^{–1} (566 stations). Details of the GPS analysis and processing techniques are included in the Methods.

Figure 1 shows the distribution of vertical GPS rates across central California and east of the Sierra Nevada into the Great Basin. Long-term rates exclude GPS stations located within the San Joaquin Valley groundwater basin that display comparatively large signals owing to their position on soft sediment. These stations are affected by compaction-driven subsidence through loss of pore water⁵, poro-elastic deformation related to groundwater level fluctuations²⁶, local irrigation and other processes unrelated to solid earth motion.

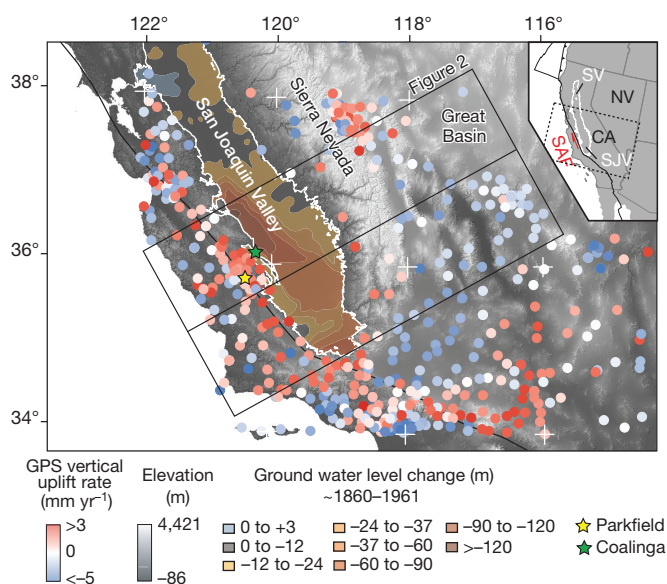


Figure 1 | Contemporary GPS vertical rates and groundwater decline. Map of vertical uplift rates from GPS stations (circles) spanning California and the western Great Basin. Stations in the valley showing anomalously larger signals and local irrigation effects are excluded. Contours show historical changes in the deep, confined aquifer¹. The inset depicts the tectonic configuration of the Central Valley groundwater basin (SV, Sacramento Valley; SJV, San Joaquin Valley), and the San Andreas Fault (SAF).

¹Geology Department, Western Washington University, Bellingham, Washington 98225-9080, USA. ²Department of Earth Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. ³Nevada Geodetic Laboratory, Nevada Bureau of Mines and Geology and Nevada Seismological Laboratory, University of Nevada, Reno, Nevada 89557, USA. ⁴Berkeley Seismological Laboratory, University of California, Berkeley, California 94720-4760, USA. ⁵Department of Earth and Planetary Science, University of California, Berkeley, California 94720-4767, USA.

Vertical rates across a transect from the central Coast Ranges to the relatively stable Great Basin range up to about 3 mm yr^{-1} . The highest average velocities concentrate along the margins of the valley (Fig. 2), near the locus of the greatest historical changes in the deep confined aquifer¹ (Fig. 1). The uplift rates of stations west of the San Joaquin Valley are more variable than those at bedrock sites in the Sierra Nevada, reflecting a higher proportion of stations on or near agricultural basins or active faults in the central Coast Ranges. Scatter in the vertical rate transect probably reflects shifts in the locus of groundwater change in comparison with historical averages², as well as some aleatory variability. Nearby earthquakes, such as the San Simeon rupture in 2003 of moment magnitude 6.5, may also influence stations near the coast with anomalously high rates near 2 mm yr^{-1} (Fig. 2). In the Sierra Nevada, the highest vertical rates occur along the western slope and steadily decay to the northeast in the Great Basin (Fig. 2). Previous studies show a similarly poor or inverse correlation between elevation and GPS uplift rates across the Sierra Nevada^{7–9}. Modelling of postseismic viscoelastic relaxation following nearby historical earthquakes suggests that such transient strain accounts for only a fraction of the total observed vertical signal in the southern Sierra Nevada⁹.

We compare vertical GPS rates with surface uplift predicted from an elastic half-space model simulating the response to load changes driven by variations in total water storage (Fig. 2) (see Methods). The model accounts for surface and subsurface deformation induced by line loads distributed across a 60-km-wide strip over the surface of an elastic half-space representing the San Joaquin Valley, the site of long-term, historical groundwater unloading. Using a range of elastic parameters, we fitted the GPS-derived vertical velocities with a rate of unloading of $8.8 (\pm 1.3) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$ (Fig. 2). We compare this estimate with the current average unloading rate in the valley based on changes in total

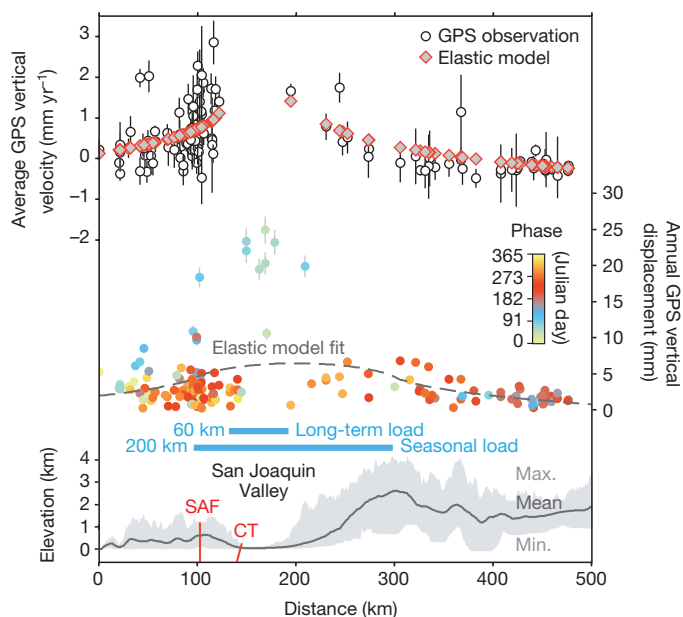


Figure 2 | GPS and model comparison. Swath profile of average contemporary vertical GPS velocity, annual GPS vertical displacement amplitude, and average topography from the central California Coast Range to the western Great Basin (SAF, San Andreas Fault; CT, Coalinga thrust). Vertical velocity and displacement amplitude are shown with 1σ uncertainties. The profile includes data from 121 stations and encompasses areas of the greatest historical and current change to groundwater levels (boxed in Fig. 1). The average GPS velocity is well fitted by an elastic model simulating surface uplift resulting from the decline in total water storage (including groundwater loss) centred along and parallel to the San Joaquin Valley. Seasonal changes in the annual GPS displacement (peak-to-peak amplitude) are distributed more broadly over the San Joaquin drainage basin, reflecting distribution of winter precipitation, snow load and reservoirs. Blue bars show the width of both the long-term and the seasonal loads used in the elastic model.

water storage, measured using satellite gravimetry from the Gravity Recovery and Climate Experiment (GRACE) (ref. 3; see Methods). This change, measured between October 2003 and March 2010, yields an average equivalent unloading rate of $7.2 (\pm 2.1) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$, slightly lower but in good overall agreement with the GPS-derived estimate.

Assuming that changes in total water storage drive vertical motion surrounding the San Joaquin Valley, the GPS data reflect a combination of short-term, elastic response to ongoing groundwater depletion and longer-term viscous relaxation reflecting the history of hydrospheric mass changes. Slight overestimation of the unloading rate based on the GPS results leaves open the possibility of either a time-dependent (viscous) component of the uplift signal, or a (smaller) contribution of ongoing tectonic uplift. Both the overall match between the elastic model and observed GPS uplift (Fig. 2) and seasonal patterns inherent in the phase and amplitude of the GPS vertical time series, however, demonstrate the importance of the instantaneous (elastic) response to groundwater unloading.

Annual peak uplift for GPS stations in the Coast Ranges and Sierra Nevada occurs in the late summer and early autumn (Fig. 3), corresponding with diminished snow and surface water loads²⁷ and overlapping with the end of the summer growing season and peak groundwater pumping in the Central Valley (May to September; ref. 2). In contrast, recharge during the early spring drives larger-amplitude peak uplift within the valley through poro-elastic effects of aquifer water levels (Fig. 3 and ref. 2). Seasonal vertical displacements are more broadly distributed than average, long-term trends (Fig. 2), reflecting the variability in total water storage in upland catchments feeding the San Joaquin Valley. Notably, stations along the valley margins move upward annually in accordance with the Sierra Nevada and Coast Ranges during dry months and show a corresponding subsidence during the wet winter and spring (Fig. 3). These peripheral stations lead to the bimodal distribution for peak uplift times observed for the Central Valley (inset to Fig. 3), indicating that uplift from hydrospheric load changes dominates even the shortest-term signals in the vertical GPS measurements for stations unaffected by local irrigation or aquifer effects.

Flexure in the central Coast Ranges due to seasonal hydrospheric unloading in the San Joaquin Valley provides a viable mechanism to explain the annual modulation of seismicity on the San Andreas Fault. Both the locked and creeping fault sections at Parkfield (Fig. 1) exhibit an increase in the number of earthquakes greater than magnitude 1.25 during the late summer and autumn (Fig. 3), previously attributed to local changes in effective stress linked to the hydrologic cycle⁶. We explore the potential seasonal impacts of unloading and uplift of the Coast Ranges on short-term changes in the fault-normal stress resolved across the fault using the elastic half-space model (see Methods). Taking estimates of annual variations of 100–300 mm in equivalent water height in the San Joaquin Valley from GRACE data³ and a wider load encompassing the Coast Ranges and the western Sierra Nevada, the elastic model produces between 3 mm and 8 mm of maximum annual vertical ground motion, in good agreement with the peak-to-peak amplitudes of annual uplift measured by GPS (Fig. 2). The corresponding fault-normal stress variations on the San Andreas Fault are near 1 kPa at seismogenic depths (Extended Data Fig. 4), with peak unclamping during the dry summer and autumn. Although we cannot rule out the potential feedback of reduced effective normal stress due to diffusion of pore fluids into fault zone rocks during recharge⁶, our results suggest that unloading may contribute to seasonal modulation of seismic activity on the central San Andreas Fault. A similar mechanism was invoked to explain annual variations in seismicity in the Himalayas¹⁸. Stress changes due to groundwater unloading are somewhat higher for other historically active faults closer to the valley such as the Coalinga thrust system (Fig. 1), where peak-to-peak annual unloading gives rise to positive Coulomb stress changes of 1.0 kPa to 1.7 kPa (see Methods and Extended Data Fig. 4).

Given that relatively small annual stress perturbations (around 1 kPa) appear sufficient to modulate earthquakes along the central San Andreas Fault⁶, we use the elastic model to estimate the stress rate caused by

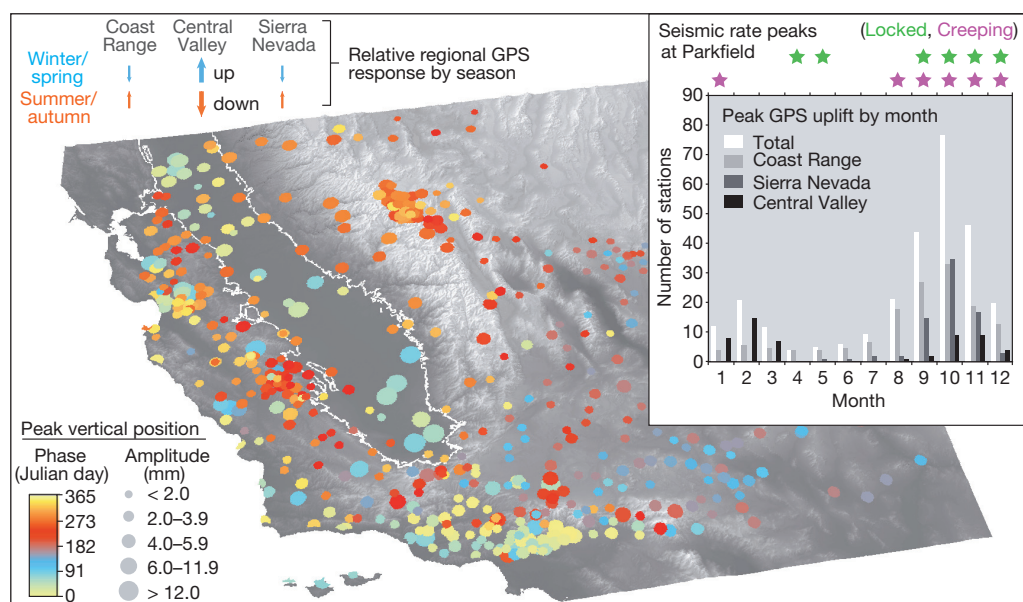


Figure 3 | Seasonal peak uplift from GPS. Phase and peak-to-peak amplitude of annual vertical GPS displacement for all stations included in our analysis. The inset shows histograms of peak uplift phase (binned by month) for stations in the San Joaquin Valley, Sierra Nevada, and Coast Range. Peaks in seismicity rate for the locked and creeping San Andreas Fault at Parkfield are defined as months with higher than average declustered seismicity⁶. Smaller-amplitude peak uplift in the Sierra Nevada and Coast Range during

the dry summer and autumn is largely out of phase with larger-amplitude uplift in the San Joaquin Valley driven by the poro-elastic response to recharge during the wetter winter and spring. Stations peripheral to the valley show uplift patterns in accordance with nearby sites on bedrock, indicating the dominance of surface loads in driving vertical motions away from areas affected by local irrigation and groundwater fluctuations.

ongoing human-induced groundwater removal (see Methods). Taking an estimate of the unloading rate due to loss in water storage in the San Joaquin Valley of $8.8 (\pm 1.3) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$, we calculate a rate of unclamping of $0.07\text{--}0.18 \text{ kPa yr}^{-1}$ for the San Andreas Fault. Coulomb failure stress rates on the Coalinga thrusts are higher, averaging $0.13\text{--}0.57 \text{ kPa yr}^{-1}$. Estimates of total groundwater loss from historical and more recent water level records (around 160 km^3 ; refs 3 and 4) correspond to a reduction in the effective normal stress of $2.7\text{--}9.5 \text{ kPa}$ for the San Andreas Fault, and a Coulomb stress change of $10\text{--}15 \text{ kPa}$ at Coalinga, since the beginning of groundwater extraction in 1860. These results suggest that human activity may give rise to a gradual increase in the rate of earthquake occurrence, as suggested by earthquake catalogues in central California⁶.

Future scenarios for groundwater in California suggest increasing demand for agricultural, urban and environmental use². Climate change will probably exacerbate the stress on this resource through altered precipitation patterns, more frequent droughts, earlier snowmelt, larger floods, and increasing temperatures and evapotranspiration^{28,29}. We demonstrate how long-term and seasonal hydrospheric mass changes due to groundwater removal significantly affect regional crustal deformation, which can lead to annual variations in small-earthquake frequency and longer-term changes of stresses on the San Andreas Fault. Our model explains recent vertical uplift rates previously attributed to epeirogenic deformation in central California¹⁰, indicating that much of this signal results from anthropogenic groundwater removal. Taken together, our results highlight new and underappreciated links between human activity and solid earth processes driven by hydrospheric change.

METHODS SUMMARY

Elastic model. A two-dimensional model accounts for deformation from a normal line load distributed over a finite-width strip on the surface of the half-space^{13,30}. The vertical displacement rate \dot{u} at the surface ($z = 0$) along a profile perpendicular to the load centred at $x = x_d$ is given by:

$$\dot{u}(x, z=0) = \frac{(1-\nu)\dot{N}_0}{\pi G(2a)} [2a + (x-x_d-a) \ln|x-x_d-a| - (x-x_d+a) \ln|x-x_d+a|] + K$$

where x and z are the horizontal and vertical dimensions, respectively, \dot{N}_0 is the rate of line load, ν is Poisson's ratio, G is the shear modulus, a is the strip half-width and K is a far-field rate offset correction. The load is centred on the middle of the valley and covers its entire width ($2a = 60 \text{ km}$). Given a range of elastic parameters ($\nu = 0.25$, $G = 35 \pm 5 \text{ GPa}$), we fit observed vertical velocities using a line-load rate with $\dot{N}_0 = 8.8 (\pm 1.3) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$ (Fig. 2).

Unloading from groundwater loss. The Sacramento and San Joaquin basins lost an estimated volume of $30.9 (\pm 2.6) \text{ km}^3$ in total water storage between October 2003 and March 2010 (ref. 3). Total water storage change in the San Joaquin Valley ($21.5 \pm 6.2 \text{ km}^3$) equates to an unloading of $3.3 (\pm 1.0) \times 10^{13} \text{ N yr}^{-1}$ over a valley length of 450 km , giving an equivalent line load of $\dot{N}_0 = 7.2 (\pm 2.1) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$, in agreement with the elastic model.

Unclamping of the San Andreas Fault. The unclamping rate (change in fault normal stress) on a vertical fault is calculated as³⁰:

$$\dot{\tau}_{xx} = \frac{\dot{N}_0}{2\pi a} [(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2) \cos(\theta_1 + \theta_2)]$$

where θ_1 and θ_2 are the angles from the load edges to any point at depth, measured clockwise from the positive x direction. For the San Andreas Fault, located about 70 km from the 60-km -wide load centre, we estimate an unclamping rate of $0.07\text{--}0.18 \text{ kPa yr}^{-1}$ at seismogenic depths ($5\text{--}15 \text{ km}$) from groundwater unloading.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2013; accepted 30 March 2014.

Published online 14 May 2014.

- Williamson, A. K., Prudic, D. E. & Swain, L. A. Ground-water flow in the Central Valley, California. *US Geol. Surv. Prof. Pap.* 1401-D (1989).
- Faunt, C. C. (ed.) Groundwater availability of the Central Valley aquifer, California. *US Geol. Surv. Prof. Pap.* 1766 (2009).
- Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett.* **38**, L03403 (2011).
- Scanlon, B. R. *et al.* Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proc. Natl Acad. Sci. USA* **109**, 9320–9325 (2012).
- Poland, J. F., Lofgren, B. E., Ireland, R. L. & Pugh, R. G. Land subsidence in the San Joaquin Valley, California, as of 1972. *US Geol. Surv. Prof. Pap.* 437-H (1975).
- Christiansen, L. B., Hurwitz, S. & Ingebritsen, S. E. Annual modulation of seismicity along the San Andreas Fault near Parkfield, CA. *Geophys. Res. Lett.* **34**, L04306 (2007).

7. Fay, N. P., Bennett, R. A. & Hreinsdóttir, S. Contemporary vertical velocity of the central Basin and Range and uplift of the southern Sierra Nevada. *Geophys. Res. Lett.* **35**, L20309 (2008).
8. Bennett, R. A., Fay, N. P., Hreinsdóttir, S., Chase, C. & Zandt, G. Increasing long-wavelength relief across the southeastern flank of the Sierra Nevada, California. *Earth Planet. Sci. Lett.* **287**, 255–264 (2009).
9. Hammond, W. C., Blewitt, G., Li, Z., Plag, H. P. & Kreemer, C. Contemporary uplift of the Sierra Nevada, western United States, from GPS and InSAR measurements. *Geology* **40**, 667–670 (2012).
10. Saleeby, J., Saleeby, Z. & Le Pourhiet, L. Epeirogenic transients related to mantle lithosphere removal in the southern Sierra Nevada region, California: Part II. Implications of rock uplift and basin subsidence relations. *Geosphere* **9**, 394–425 (2013).
11. Holzer, T. L. Elastic expansion of the lithosphere caused by groundwater depletion. *J. Geophys. Res.* **84**, 4689–4698 (1979).
12. van Dam, T. *et al.* Crustal displacements due to continental water loading. *Geophys. Res. Lett.* **28**, 651–654 (2001).
13. Jiang, Y., Dixon, T. H. & Wdowinski, S. Accelerating uplift in the North Atlantic region as an indicator of ice loss. *Nature Geosci.* **3**, 404–407 (2010).
14. Nof, R. N. *et al.* Rising of the lowest place on Earth due to Dead Sea water-level drop: evidence from SAR interferometry and GPS. *J. Geophys. Res.* **117**, B05412 (2012).
15. Fu, Y. & Freymueller, J. T. Seasonal and long-term vertical deformation in the Nepal Himalaya constrained by GPS and GRACE measurements. *J. Geophys. Res.* **117**, B03407 (2012).
16. Heki, K. Snow load and seasonal variation of earthquake occurrence in Japan. *Earth Planet. Sci. Lett.* **207**, 159–164 (2003).
17. Luttrell, K., Sandwell, D., Smith-Konter, B., Bills, B. & Bock, Y. Modulation of the earthquake cycle at the southern San Andreas fault by lake loading. *J. Geophys. Res.* **112**, B08411 (2007).
18. Bettinelli, P. *et al.* Seasonal variations of seismicity and geodetic strain in the Himalaya induced by surface hydrology. *Earth Planet. Sci. Lett.* **266**, 332–344 (2008).
19. González, P. J., Tiampo, K. F., Palano, M., Cannavó, F. & Fernández, J. The 2011 Lorca earthquake slip distribution controlled by groundwater crustal unloading. *Nature Geosci.* **5**, 821–825 (2012).
20. Ben-Zion, Y. & Allam, A. A. Seasonal thermoelastic strain and postseismic effects in Parkfield borehole dilatometers. *Earth Planet. Sci. Lett.* **379**, 120–126 (2013).
21. Christensen, M. N. Late Cenozoic crustal movements in the Sierra Nevada of California. *Geol. Soc. Am. Bull.* **77**, 163–182 (1966).
22. Small, E. E. & Anderson, R. S. Geomorphically driven late Cenozoic rock uplift in the Sierra Nevada, California. *Science* **270**, 277–281 (1995).
23. Le Pourhiet, L. & Saleeby, J. Lithospheric convective instability could induce creep along part of the San Andreas fault. *Geology* **41**, 999–1002 (2013).
24. Blewitt, G., Kreemer, C., Hammond, W. C. & Goldfarb, J. M. Terrestrial reference frame NA12 for crustal deformation studies in North America. *J. Geodyn.* **72**, 11–24 (2013).
25. Altamimi, Z., Collilieux, X. & Métivier, L. ITRF2008: an improved solution of the international terrestrial reference frame. *J. Geod.* **85**, 457–473 (2011).
26. Schmidt, D. A. & Bürgmann, R. Time-dependent land uplift and subsidence in the Santa Clara valley, California, from a large interferometric synthetic aperture radar data set. *J. Geophys. Res.* **108**, 2416 (2003).
27. Argus, D., Fu, Y. & Landerer, F. Seasonal variation in total water storage in California inferred from GPS observations of vertical land motion. *Geophys. Res. Lett.* **41**, 1971–1980 (2014).
28. Dettinger, M. D. & Cayan, D. R. Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California. *J. Clim.* **8**, 606–623 (1995).
29. California Department of Water Resources. California water plan update 2005. Volume 1: Strategic Plan. *Dep. Water Res. Bull.* 160–05 (2005).
30. Jaeger, J., Cook, N. & Zimmerman, R. *Fundamentals of Rock Mechanics* 4th edn (Blackwell, 2007).

Acknowledgements Funding for this work comes from NSF EarthScope award number EAR-1252210 to G.B. and W.C.H. GPS data were collected using the EarthScope Plate Boundary Observatory, SCIGN, BARGEN, BARD, CORS and IGS networks. We are particularly grateful to UNAVCO for operating the vast majority of GPS stations used in this project. GPS data were processed using the GIPSY OASIS II software and data products from the Jet Propulsion Laboratory.

Author Contributions C.B.A. and P.A. performed the analysis and wrote the paper. W.C.H. and G.B. analysed and processed the GPS data. All authors contributed to the interpretations and preparation of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B.A. (colin.amos@u.washington.edu).

METHODS

GPS data processing. We use data from continuously recording GPS stations between -122° and -114° longitude and 34° to 38° latitude that had time series longer than 2.5 years that were $>50\%$ complete (Extended Data Fig. 1). This data set includes 566 continuous GPS stations within this geographic region. No episodic campaign or semi-continuous stations were included. Percentage completeness is defined as the number of daily solutions for the station divided by the number of potential daily solutions given the duration of the time series. Larger values for percentage completeness are preferred in order to best resolve seasonal oscillations of position.

The data were processed as a part of a $>11,000$ station mega-network analysis system that retrieves data daily and updates solutions weekly. We use the GIPSY/OASIS software provided by the Jet Propulsion Laboratory to estimate station coordinates every 24 h using the Precise Point Positioning method³¹. Ionosphere-free combinations of carrier phase and pseudorange were obtained every 5 min. Calibrations were applied for all antennas, ground receivers and satellite transmitters. To model tropospheric refractivity, the Global Mapping Function was applied³², with tropospheric wet zenith delay and horizontal gradients estimated as stochastic random-walk parameters every 5 min (ref. 33). The observable model includes ocean tidal loading (including companion tides) coefficients supplied by Chalmers University³⁴. Ambiguity resolution was applied to double differences of the estimated one-way bias parameters³⁵, using the wide lane and phase bias method, which phase-connects individual stations to IGS stations in common view³⁶. Satellite orbit and clock parameters were provided by the Jet Propulsion Laboratory, who determine these parameters in a global fiducial-free analysis using a subset of the available IGS core stations as tracking sites. Output station coordinates are initially in the loose frame of the Jet Propulsion Laboratory's fiducial-free GPS orbits. These were transformed into reference frame IGS08 using daily seven-parameter transformations supplied by the Jet Propulsion Laboratory. IGS08 is derived from ITRF2008 (ref. 37) and consists of 232 globally distributed IGS stations³⁸. Finally, the solutions are aligned with our custom reference frame (NA12) that co-rotates with stable North America²⁴. This alignment is applied to each daily solution and provides a spatial filter to suppress errors correlated at the continental scale. The mean formal uncertainty in daily vertical coordinate is <3.0 mm. A summary reference table for GPS solution methodology is given at <http://geodesy.unr.edu/gps/ngl.acn>. The resulting time series are used to estimate rates of motion with respect to stable North America.

GPS time series analysis. An outlier exclusion criterion was applied to remove daily solutions that were more than 20 mm from the expected position based on a provisional time series model, or having position uncertainty greater than 20 mm. For each of the three components (east, north, up) of each GPS time series we solved for the intercept b , rate v , steps at the times $t_{\text{eqp},i}$ of N known equipment changes that can introduce discontinuities of size D . We solve for annual and semiannual seasonal terms by solving for the cosine (C_i) and sine (S_i) terms. For example, the east component is:

$$e(t) = b + vt + \sum_{i=1}^N D_i \mathcal{H}(t - t_{\text{eqp},i}) + \sum_{k=1}^2 [C_k \cos(2\pi kt) + S_k \sin(2\pi kt)]$$

where $\mathcal{H}(t)$ is the Heaviside step function, where t is in years. We use a damped linear inversion with weights on the prior uncertainties of the model parameters scaled to allow the expected range of values for each parameter. However, because only continuously recording GPS stations were used, all parameters in the estimation (including seasonal terms) were well constrained, so the damping parameters had little influence on the solutions. The amplitude of the annual terms were calculated with:

$$A = \sqrt{C^2 + S^2}$$

and the day of year from:

$$\text{day} = \left(\frac{365}{2\pi} \right) \tan^{-1} \left(\frac{S}{C} \right)$$

Finally, stations were excluded if their rate uncertainty was greater than 1 mm yr^{-1} or their rate was clearly anomalous (that is, greater than 100 mm yr^{-1}). In all, we estimated solutions for 566 stations, 96% of which have time series duration greater than 4.0 years, and 92% of which are 80% or more complete (Extended Data Fig. 2). Example vertical GPS time series are shown in Extended Data Fig. 3. GPS time series parameters, including vertical uplift rate and the amplitude and timing of peak uplift are included as source data for Fig. 1.

Elastic half-space model. We use an elastic half-space model for calculating the elastic response of the crust to changes in total water storage. The model is two-dimensional and accounts for surface and subsurface deformation induced by a line load over the surface of the half-space along a finite-width strip (half-width a ,

Extended Data Fig. 4a). Model derivation can be found in ref. 30, with a small correction from ref. 13. The normal displacement of the surface ($z = 0$) along a profile perpendicular to the load centred at $x = x_d$ is given by

$$u(x, z=0) = \frac{(1-\nu)N_0}{\pi G(2a)} [2a + (x - x_d - a) \ln|x - x_d - a| - (x - x_d + a) \ln|x - x_d + a|] + K$$

where x and z are the horizontal and vertical dimensions, respectively, N_0 is a line load, ν is Poisson's ratio, G is the shear modulus, a is the strip half-width and K is a far-field rate offset correction. We fix the position of the line load to the centre of the San Joaquin Valley, and use a load width of $2a = 60 \text{ km}$ corresponding to the average width of the valley floor. Given a range of elastic parameters ($\nu = 0.25$, $G = 35 \pm 5 \text{ GPa}$), we can fit the vertical velocities derived from the GPS data by a line-load rate with $\dot{N}_0 = 8.8 (\pm 1.3) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$ (see Fig. 2). Goodness of fit is estimated using a reduced chi-square criterion. Our best-fit model has a reduced chi-square of 1.9, indicating a reasonably good fit.

Unloading from estimated mass loss. The combined Sacramento–San Joaquin River basins lost an estimated volume of $30.9 (\pm 2.6) \text{ km}^3$ in total water storage over a 6.5-year period between October 2003 and March 2010 (ref. 3). Of this volume, $20.3 (\pm 3.8) \text{ km}^3$ is attributed to loss of groundwater in the Central Valley, 80% of which occurred in the San Joaquin Valley. Assuming that changes in total water storage not related to groundwater (for example, soil moisture, surface water and snow) are distributed equally between the Sacramento and San Joaquin basins, the total water storage change in the San Joaquin Valley over the 6.5-year period is $21.5 (\pm 6.2) \text{ km}^3$. This change is equivalent to a rate of mass loss of $3.3 (\pm 1.0) \times 10^3$ gigatons per year, and therefore a rate of unloading of $3.2 (\pm 1.0) \times 10^{13} \text{ N yr}^{-1}$ distributed over the 450 km length of the valley. Taking into account the range of uncertainties in groundwater and total water storage change³, the equivalent line-load rate is thus $\dot{N}_0 = 7.2 (\pm 2.1) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$, in agreement with the estimate from the elastic model based on the vertical GPS velocities. If we add the estimated 1860–2003 depletion of 140 km^3 (ref. 4) to the 2003–2010 water loss, the cumulative historic load decrease amounts to $\sim 3.5 \times 10^9 \text{ N m}^{-1}$.

Stress modelling. The two-dimensional stress components at any point at depth caused by a distributed line load N_0 (in N m^{-1} , the negative value indicates unloading) at the surface of the elastic half-space are calculated as³⁰:

$$\tau_{xx} = \frac{N_0}{2\pi a} [(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2) \cos(\theta_1 + \theta_2)]$$

$$\tau_{zz} = \frac{N_0}{2\pi a} [(\theta_1 - \theta_2) - \sin(\theta_1 - \theta_2) \cos(\theta_1 + \theta_2)]$$

$$\tau_{xz} = \frac{N_0}{2\pi a} [\sin(\theta_1 - \theta_2) \sin(\theta_1 + \theta_2)]$$

where θ_1 and θ_2 are the angles from both edges of the load measured clockwise from the positive x direction, a is the load half-width and z is positive downward (Extended Data Fig. 4a). The shear and normal stress components resolved on a fault plane dipping at an angle φ with a strike direction normal to the xz plane are given by:

$$\tau_n = \tau_{zz} \cos^2(\varphi) - 2\tau_{xz} \sin(\varphi) \cos(\varphi) + \tau_{xx} \sin^2(\varphi)$$

$$\tau_z = (\tau_{zz} - \tau_{xx}) \sin(\varphi) \cos(\varphi) + \tau_{xz} (\cos^2(\varphi) - \sin^2(\varphi))$$

Using Coulomb failure assumptions³⁰, the Coulomb failure stress is expressed as:

$$\sigma_c = |\tau_z| + \mu(\tau_n - p) + S$$

where μ is the coefficient of friction, p is pore-fluid pressure and S is cohesion. Assuming that p , S and μ are constant over time, the change in Coulomb stress is given by³⁹:

$$\Delta\sigma_c = \Delta|\tau_z| + \mu\Delta\tau_n$$

where $\Delta\tau_n$ is positive for tension^{30,39}. Similarly, the Coulomb stressing rate is given by:

$$\dot{\sigma}_c = |\dot{\tau}_z| + \mu\dot{\tau}_n$$

The first term on the right-hand side is valid for an isotropic failure plane. We consider the change in shear stress in the slip direction (or rake) and vary the sign

of the change in shear stress accordingly. For the San Andreas Fault we consider a dip angle of 90° and the unclamping stress (or stress rate) is given simply by $\Delta\tau_{xx}$ (or $\dot{\tau}_{xx}$). For the Coalinga blind thrust we use dip angles ranging between 15° and 30° (refs 40 and 41) and a negative change in shear stress. We also consider a 65° northeast-dipping fault with a positive slip direction, simulating stress conditions on the Nuñez thrust fault. For each fault we calculate the normal and shear stress components at depths of 5 km and 15 km, representative of the seismogenic depth in Central California.

Rate of unclamping of the San Andreas Fault. The rate of unclamping on a vertical fault (change in fault normal stress) can be calculated as³⁰:

$$\dot{\tau}_{xx} = \frac{\dot{N}_0}{2\pi a} [(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2) \cos(\theta_1 + \theta_2)]$$

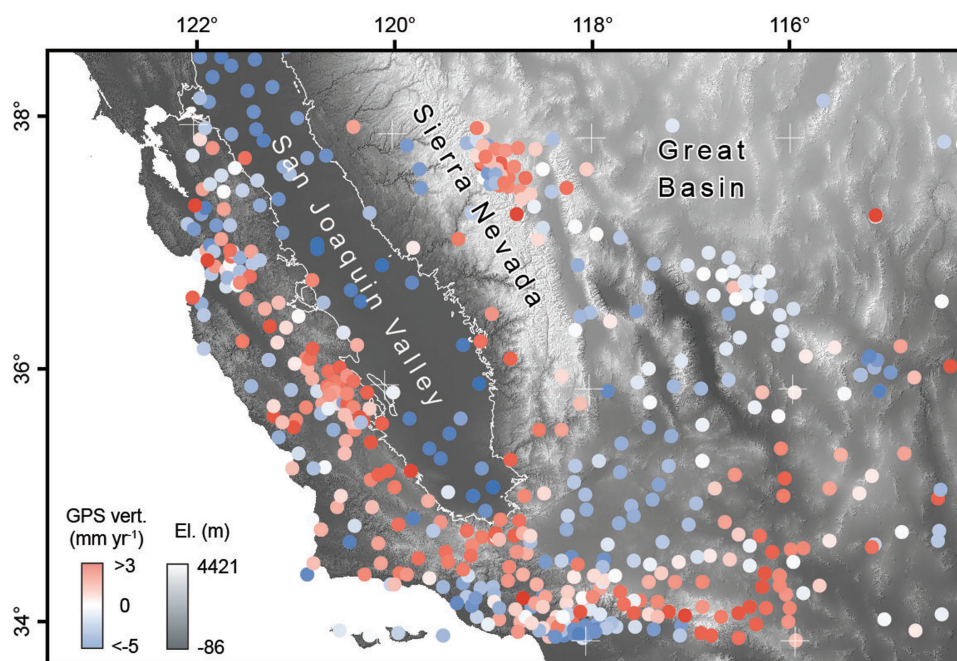
where θ_1 and θ_2 are the angles from both edges of the loading strip to any point at depth, measured clockwise from the positive x direction. Taking an estimate of the unloading rate due to loss in water storage of $8.8 (\pm 1.3) \times 10^7 \text{ N m}^{-1} \text{ yr}^{-1}$, as well as average spatial parameters for the San Andreas Fault (load width of $2a = 60 \text{ km}$; distance to centre of the distributed line load of around 70 km ; seismogenic depths of $5\text{--}15 \text{ km}$), we estimate a rate of unclamping of $0.07\text{--}0.18 \text{ kPa yr}^{-1}$ at the seismogenic depth along the San Andreas Fault due to the elastic rebound caused by unloading (Extended Data Fig. 4b and c).

For the Coalinga blind thrust, we calculate the normal (τ_n) and shear (τ_s) stress rates resolved on a range of fault planes and depths. The Coulomb failure stress rate (ignoring pore pressure), $\dot{\sigma}_c = |\dot{\tau}_z| + \mu \dot{\tau}_n$ (ref. 30), is then calculated with $\mu = 0.7$.

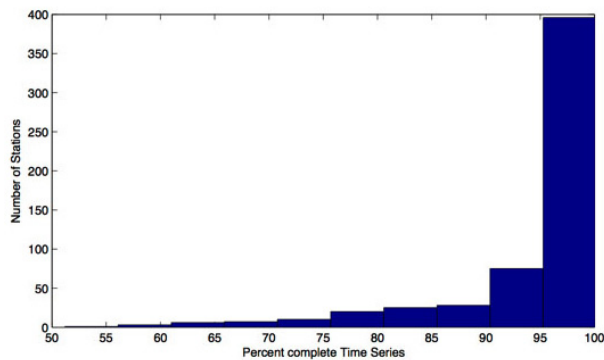
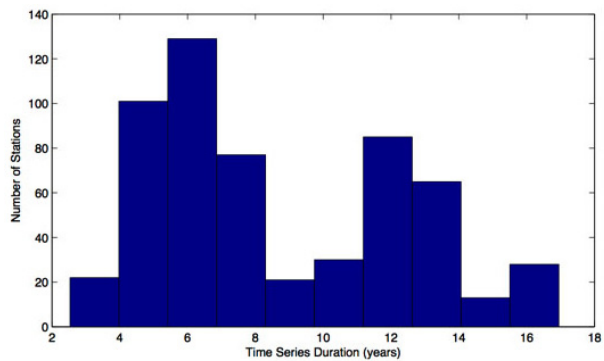
Seasonal vertical displacements and stress changes. The elastic model can be used to estimate total vertical displacements caused by variations in total water storage. The model considered is more broadly distributed over the Coast Range, the San Joaquin Valley and the western Sierra Nevada (width of 200 km ; Fig. 2), corresponding to the approximate total width of the San Joaquin River drainage basin. Taking an average of 200 mm of change in total water storage, we estimate a load of $\sim 4 \times 10^8 \text{ N m}^{-1}$, which produces annual vertical displacements of $3\text{--}8 \text{ mm}$ within

the load extent (Fig. 2). The associated stress variations on the San Andreas Fault are about 1 kPa and the Coulomb stress changes on the Coalinga thrust are $1.0\text{--}1.7 \text{ kPa}$ (Extended Data Fig. 4d and e). Varying the width of the distributed line loads by $\pm 20 \text{ km}$ and the load centre by $\pm 10 \text{ km}$ has only a small impact ($< 0.5 \text{ kPa}$) on the resolved seasonal stress changes on the faults.

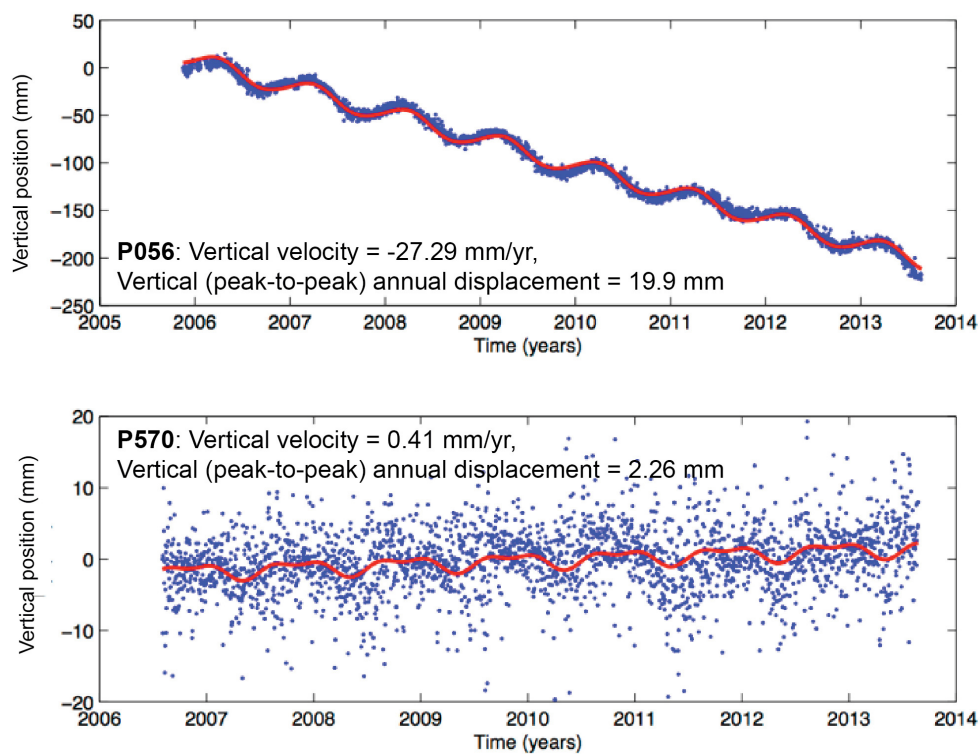
31. Zumberge, J. F., Heflin, M. B., Jefferson, D. C., Watkins, M. M. & Webb, F. H. Precise point positioning for the efficient and robust analysis of GPS data from large networks. *J. Geophys. Res.* **102**, 5005–5017 (1997).
32. Boehm, J., Niell, A., Tregoning, P. & Schuh, H. Global mapping function (GMF): a new empirical mapping function based on numerical weather model data. *Geophys. Res. Lett.* **33**, L07304 (2006).
33. Bar-Sever, Y. E., Kroger, P. M. & Borjesson, J. A. Estimating horizontal gradients of tropospheric path delay with a single GPS receiver. *J. Geophys. Res.* **103**, 5019–5035 (1998).
34. Scherneck, H.-G. A parametrized solid earth tide model and ocean tide loading effects for global geodetic baseline measurements. *Geophys. J. Int.* **106**, 677–694 (1991).
35. Blewitt, G. Carrier phase ambiguity resolution for the Global Positioning System applied to geodetic baselines up to 2000 km . *J. Geophys. Res.* **94**, 10187–10283 (1989).
36. Bertiger, W. *et al.* Single receiver phase ambiguity resolution with GPS data. *J. Geod.* **84**, 327–337 (2010).
37. Altamimi, Z., Collilieux, X. & Métivier, L. ITRF2008: an improved solution of the international terrestrial reference frame. *J. Geodesy* **85**, 457–473 (2011).
38. Rebischung, P. *et al.* IGS08: The IGS realization of ITRF2008. *GPS Solut.* **16**, 483–494 (2012).
39. Harris, R. A. Stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res.* **103**, 24347–24358 (1998).
40. Stein, R. S. & Ekstrom, G. E. Seismicity and geometry of a 110-km long blind thrust fault. 2. Synthesis of the 1982–1985 California earthquake sequence. *J. Geophys. Res.* **97**, 4865–4883 (1992).
41. Toda, S. & Stein, R. S. Response of the San Andreas fault to the 1983 Coalinga–Nuñez earthquakes: an application of interaction-based probabilities. *J. Geophys. Res.* **107**, 2126 (2002).



Extended Data Figure 1 | Vertical GPS data. Map of vertical uplift rates spanning California and the western Great Basin, including stations within the Central Valley groundwater basin. El., elevation; GPS vert., GPS vertical velocity.

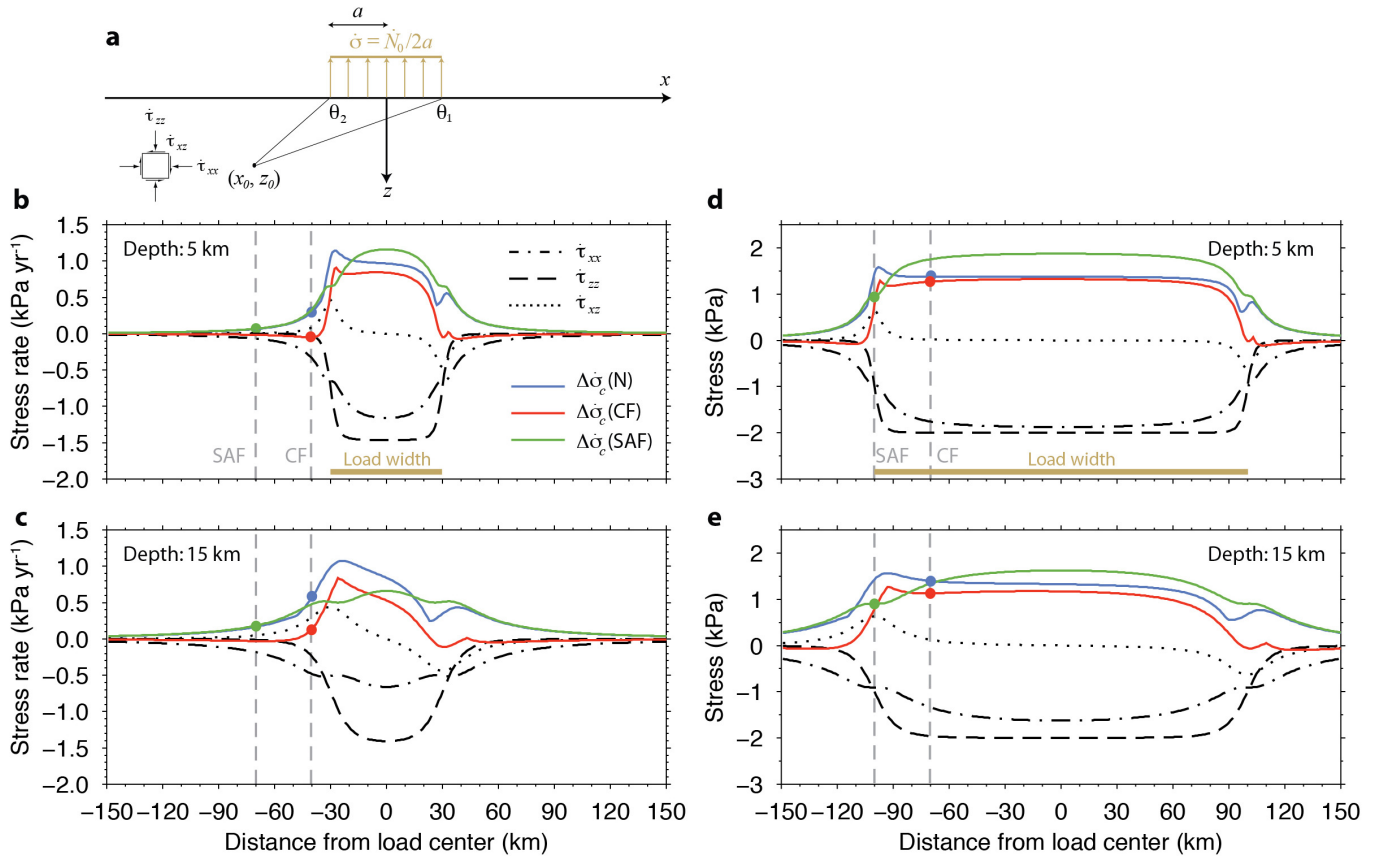


Extended Data Figure 2 | GPS time series. The top panel shows a histogram of the duration of GPS time series. 96% are longer than 4.0 years. The bottom panel shows a histogram of the percentage of complete time series. 92% are greater than 80% complete.



Extended Data Figure 3 | GPS time series. Example vertical GPS time series for stations P056 and P570. Blue dots are daily vertical position solutions, red lines are the model estimated to represent the time series. The station

P056 is in the southern Central Valley near Porterville, California. The station P570 is near Weldon, California, east of Lake Isabella.



Extended Data Figure 4 | Stress profiles from distributed line loads of an elastic half-space. **a**, Model setup, where \dot{N}_0 is the line-load rate, a is the load half-width, and θ_1 and θ_2 measure the angle downward from the positive x axis to any point (x_0, z_0) at depth. $\dot{\tau}$ and $\Delta\dot{\sigma}_c$ give the stress rates and Coulomb stress rates, respectively. **b**, **c**, Stress rates at depths of 5 km (**b**) and 15 km (**c**) calculated from the long-term rate of unloading over the San Joaquin Valley with $a = 30$ km. Vertical dashed lines labelled SAF and CF represent the locations of the San Andreas Fault and the Coalinga blind thrust faults (including the Nuñez fault), respectively, relative to the load centre. Black and coloured lines indicate stress components and Coulomb stress changes, respectively. The blue curves labelled N show Coulomb stress calculations for

favourably oriented faults with a 65° dip, representing the Nuñez fault. For the red curve representing the Coalinga fault (CF), the calculations are performed using a dip of 30° for unfavourably oriented faults. The green curve represents unclamping stress for vertically dipping faults such as the San Andreas Fault. **d**, **e**, Stress changes at depths of 5 km and 15 km from seasonal (peak-to-peak) load changes over the San Joaquin River Basin with $a = 100$ km (full width of 200 km). The position of the San Andreas Fault and Coalinga faults reflects displacement of the load centre by 30 km relative to the long-term load. Varying the load half-width and the load centre by ± 10 km has only a small impact (<0.5 kPa) on the resolved seasonal stress changes on the faults.

High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells

Yuxin Zhou^{1*}, Shiyu Zhu^{1*}, Changzu Cai^{1*}, Pengfei Yuan¹, Chunmei Li², Yanyi Huang² & Wensheng Wei¹

Targeted genome editing technologies are powerful tools for studying biology and disease, and have a broad range of research applications^{1–7}. In contrast to the rapid development of toolkits to manipulate individual genes, large-scale screening methods based on the complete loss of gene expression are only now beginning to be developed^{8,9}. Here we report the development of a focused CRISPR/Cas-based (clustered regularly interspaced short palindromic repeats/CRISPR-associated) lentiviral library in human cells and a method of gene identification based on functional screening and high-throughput sequencing analysis. Using knockout library screens, we successfully identified the host genes essential for the intoxication of cells by anthrax and diphtheria toxins, which were confirmed by functional validation. The broad application of this powerful genetic screening strategy will not only facilitate the rapid identification of genes important for bacterial toxicity but will also enable the discovery of genes that participate in other biological processes.

Recent progress in genomic editing, such as ZFN (zinc finger nuclease)^{1–3}, TALENs (transcription activator-like effector nucleases)^{3–5} and the CRISPR/Cas system^{6,7}, has markedly changed the way that researchers study genes and their functions in mammalian systems. The CRISPR/Cas system was originally found to provide bacteria and archaea with adaptive immunity against viruses and plasmids¹⁰. By means of the type II CRISPR system, the Cas9 nuclease can be directed by a chimeric single guide RNA (sgRNA) via Watson–Crick base pairing¹¹ to the desired genomic loci followed by an NGG protospacer-adjacent motif (PAM) to create double-stranded DNA breaks (DSBs)^{6,7}. Taking advantage of this effective gene-editing technology with easy programmability, we developed a focused sgRNA library that enables gene identification from function-based genetic screening.

The CRISPR/Cas system requires both the gene-specific sgRNA and an identical nuclease Cas9 to achieve DSBs at the targeted locus^{6,7}. We developed cell lines with a constant expression of Cas9 with OCT1 (Fig. 1a and Supplementary Fig. 1), as it has been reported that overexpression of OCT1 would further boost the sgRNA expression through enhanced U6 promoter activity¹². We used the Golden Gate cloning method¹³ to construct the sgRNA on the lentiviral backbone (Fig. 1b and Supplementary Fig. 1). To verify whether the expression of sgRNA through lentiviral delivery is sufficient to guide Cas9 to the targeted loci to generate DSBs, we arbitrarily selected the *CSPG4* gene for targeting and conducted the T7E1 digestion assay as previously described¹⁴. *CSPG4*-specific sgRNA was found to induce insertions and deletions (indels) through viral infection in all three lines constantly expressing Cas9 and OCT1 (that is, HEK293T_{OC}, HT1080_{OC} and HeLa_{OC}, where OC indicates OCT1 and Cas9). However, the pooled HeLa_{OC} cells showed lower efficiency in creating indels than the pooled HEK293T_{OC} and HT1080_{OC} cells at day 5 after infection, whereas one isolated single clone of HeLa_{OC}, designated as HeLa_{OC}-SC, showed the highest level of efficiency (Fig. 1c). This difference in indel efficiency is unlikely to be due to the variable levels of Cas9 expression (Extended Data Fig. 1), and prolonged culturing enabled HeLa_{OC} at day 8 to reach similar

cleavage efficiency as HEK293T_{OC} and HT1080_{OC} at day 5 (Fig. 1c), although still significantly lower than HeLa_{OC}-SC. It is therefore critical to isolate a single clone with high DSB mis-rejoining activity, at least for certain cell types. The beneficial role of exogenous expression of OCT1 requires further investigation.

We next examined the off-target activity of virally delivered sgRNA in HeLa_{OC}-SC. Three representative sgRNAs targeting three genes (*ANTXR1*, *HBEFG* and *CSPG4*) were chosen for this analysis. The top five off-target sites were selected for each sgRNA, and the T7E1 assays indicated that there were no off-target cleavages except for one

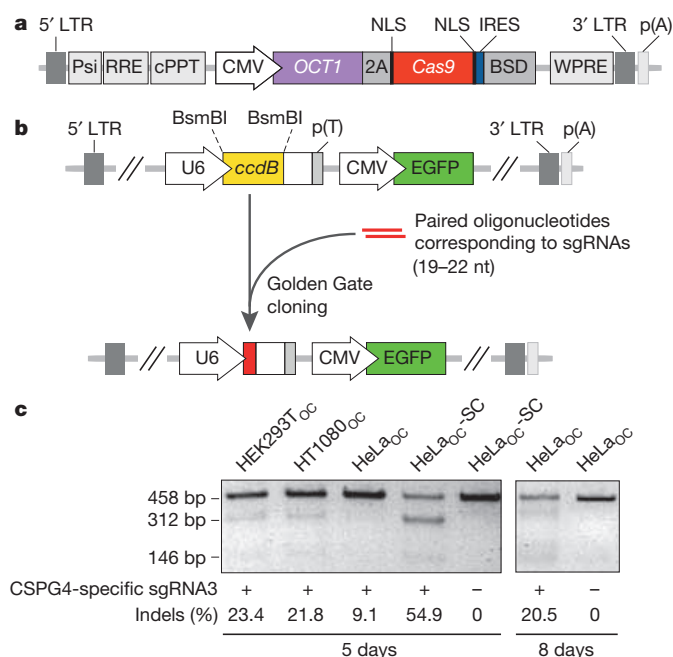


Figure 1 | Lentiviral-delivered sgRNA creates indels with high efficiency in human cells stably expressing Cas9 and OCT1. **a**, Structure of the lentiviral plasmid expressing 2A-linked OCT1 and Cas9. An IRES-linked blasticidin-resistant gene was fused with the *cas9* gene to facilitate the selection of cell clones with high-level expression of Cas9 and OCT1. The pLenti-CMV-BSD vector was used for plasmid construction. nt, nucleotides; LTR, long terminal repeat. **b**, Schematic diagram of the sgRNA expression construct in a lentiviral backbone. U6 promoter-driven *ccdB* and sgRNA scaffold were cloned into the LL3.7 lentiviral vector. Annealed oligonucleotide pairs encoding customized sgRNA were ligated into the lentiviral backbone using the Golden Gate method. **c**, Indels induced by lentivirus-delivered sgRNA (5'-TTGGCCAGACTTG CATCCG-3') targeting the *CSPG4* gene (*CSPG4*-specific sgRNA3) in the indicated cells were assayed by T7E1 digestion. All cells were incubated for 5 or 8 days after infection before assay as indicated. Primers used for the PCR reactions are listed in Supplementary Table 5, and the percentage of cleaved band was measured using the ImageJ program (<http://rsbweb.nih.gov/ij/>) for this and other figures.

¹State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China. ²Biodynamic Optical Imaging Centre (BIOPIIC), College of Engineering, Peking University, Beijing 100871, China.

*These authors contributed equally to this work.

case, in which the off-target site of sgRNA targeting *HBEGF* has exactly the same sequence for the last 15 base pairs immediately upstream of PAM (Supplementary Table 1). All of the off-target sites tested were located in the non-coding regions, and therefore unlikely to alter gene function.

To establish our methodology, we created a library targeting 291 human genes that fit our specific research interests. Assisted by a rule⁷-based algorithm, we designed three kinds of sgRNAs targeting each gene for a majority of the 291 genes (Supplementary Table 2). The single clone HeLa_{OC}-SC was chosen to produce the library because this particular line generated the highest level of indels (Fig. 1c). The library, designated as HeLa/sgrNA^{library}, was created in such a way that the coverage of sgRNAs is approximately 1,000-fold with a virus MOI (multiplicity of infection) of 0.05. We then applied this library to a functional screening process using two bacterial toxins: diphtheria and chimeric anthrax (PA/LFnDTA, protective antigen (PA)/N-terminal domain of lethal factor (LF) fused to the catalytic subunit of diphtheria toxin) toxins. Figure 2 displays a schematic of the library construction and functional screening. After three rounds of toxin treatment, the surviving cells were collected, most of which exhibited green fluorescence (Extended Data Fig. 2), indicating the active presence of sgRNA-containing cartridges that carry CMV promoter-driven EGFP (Fig. 1b). The genomic DNA of the pooled surviving cells from the library, as well as the original library of cells before the toxin treatment, was extracted and used for PCR amplification of the sgRNA-coding regions, before being subjected to deep-sequencing analysis (Fig. 2). These

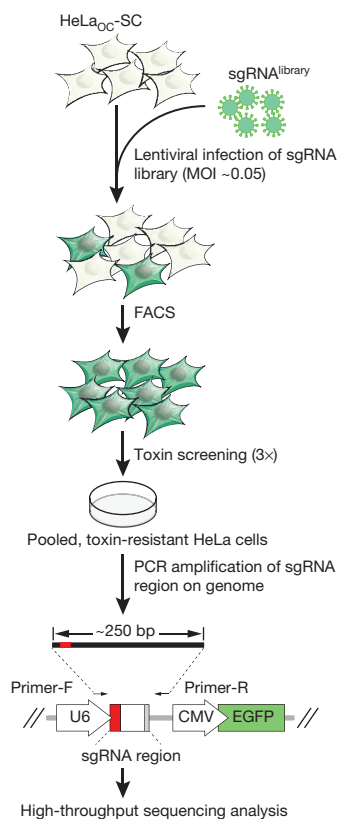


Figure 2 | Schematic of sgRNA library construction and functional screening. sgRNAs were delivered into HeLa_{OC}-SC cells by lentiviral infection with a MOI of 0.05. Library cells stably expressing sgRNAs were obtained by FACS for green fluorescence 2 days after infection. Library screening was conducted by three rounds of toxin treatment, followed by PCR amplification of the sgRNA-coding sequence integrated into the chromosomes. Primers used for the PCR reaction are listed in Supplementary Table 5. Purified PCR product was subjected to high-throughput sequencing analysis with Illumina HiSeq 2500.

sgRNA-coding regions can be amplified rather specifically without interference from the host chromosomal DNA (Extended Data Fig. 3).

High-throughput sequencing analysis revealed a total of 863 (99.3%) sgRNA sequences from the control library that was designed to contain 869 kinds of sgRNA targeting a total of 291 genes (Supplementary Table 2). The read counts for most of the sgRNA coding sequences were consistent among three repeats (Supplementary Table 3 and Extended Data Fig. 4a) and after prolonged culturing (Extended Data Fig. 4b). After three rounds of toxin treatment, however, the composition and the read count of the sgRNAs showed distinct differences between the two different screens. We organized all the log₂-fold changes of the normalized counts into two heat maps, ranked according to their enrichment level from either PA/LFnDTA or diphtheria toxin selection. Samples of the top hits in each screening were also shown with their gene names and sgRNA numbers. Most of the sgRNAs were depleted in both cases, as expected (Fig. 3a). Using the analytical tool DESeq2 (ref. 15), we selected 21 sgRNAs targeting 19 genes from PA/LFnDTA screening and 15 sgRNAs targeting 15 genes from diphtheria toxin screening as potential candidates for future studies (Figs 3b, c and Supplementary Table 4). Notably, all three sgRNAs targeting the same known anthrax toxin receptor gene *ANTXR1* (also called *TEM8*) (ref. 16) were enriched and ranked at the top from the screening of PA/LFnDTA (Fig. 3b), and one of the two sgRNAs targeting the diphtheria toxin receptor gene *HBEGF* (ref. 17) ranked at the top from the screening of the diphtheria toxin (Fig. 3c).

Notably, the three *ANTXR1*-targeting sgRNAs varied in their efficiency of inducing gene interruption, as shown by the rates of indels determined by a T7E1 digestion assay. The two most strongly enriched sgRNAs (sgRNA1 and sgRNA3) showed a much higher efficiency in causing indels at the targeted locus of the *ANTXR1* gene (Fig. 3d). Consistently, the two sgRNAs targeting *HBEGF* also had a distinct efficiency in inducing indels (Fig. 3e), with *HBEGF* sgRNA3, with a much higher level of efficiency, being enriched and ranked at the top of the diphtheria toxin screening results (Fig. 3a), whereas the other sgRNA targeting *HBEGF* ranked at number 35 (Supplementary Table 4).

To validate the functions of candidate genes identified from the library screening, we generated individual gene knockouts in HeLa cells for *ANTXR1* and *HBEGF*. The *ANTXR1* knockout in HeLa cells was created by either the TALEN technique¹⁸ or CRISPR/Cas9 system (Extended Data Fig. 5a, b). We tested all three mutant clones (one created by TALENs and the other two created by CRISPR/Cas9) for their susceptibility to the two toxins using both XTT (2,3-bis-(2-methoxy-4-nitro-5-sulphophenyl)-2H-tetrazolium-5-carboxanilide) cell viability and LDH (lactate dehydrogenase) cell death assays. The HeLa *ANTXR1*^{-/-} became totally resistant to PA-mediated toxicity, while remaining susceptible to the diphtheria toxin (Fig. 4a and Extended Data Fig. 5c, d). Similarly, HeLa *HBEGF*^{-/-} was completely resistant to diphtheria toxin, but not to PA/LFnDTA for two independent clones created through the CRISPR/Cas9 system (Fig. 4b and Extended Data Fig. 6). The identification and confirmation of these two known toxin receptors demonstrates the high level of efficiency and reliability of the CRISPR/Cas9 library in functional screening.

To determine whether our screening revealed novel genes involved in the two toxins' intoxication mechanisms, we picked candidate genes from the top of sgRNA enrichment lists (Fig. 3b, c) for further validation. Because only one sgRNA was positively selected for most of the candidates, we designed additional sgRNAs (Supplementary Table 6) and examined their capacity to produce a toxin-resistance phenotype. Four genes (*PLXNA1*, *FZD10*, *PECR* and *CD81*) were confirmed to be potential functional candidates specifically involved in PA-mediated anthrax toxicity because at least two different sgRNAs for each candidate produced a toxin-resistance phenotype (Fig. 4c and Extended Data Fig. 7). As there was only one *CFTR*-targeting sgRNA conferring resistance, further analysis was required to rule out the possibility that this is not due to an off-target effect. Similarly, the involvement of *RAB2A* in diphtheria toxin toxicity was supported by the effect of

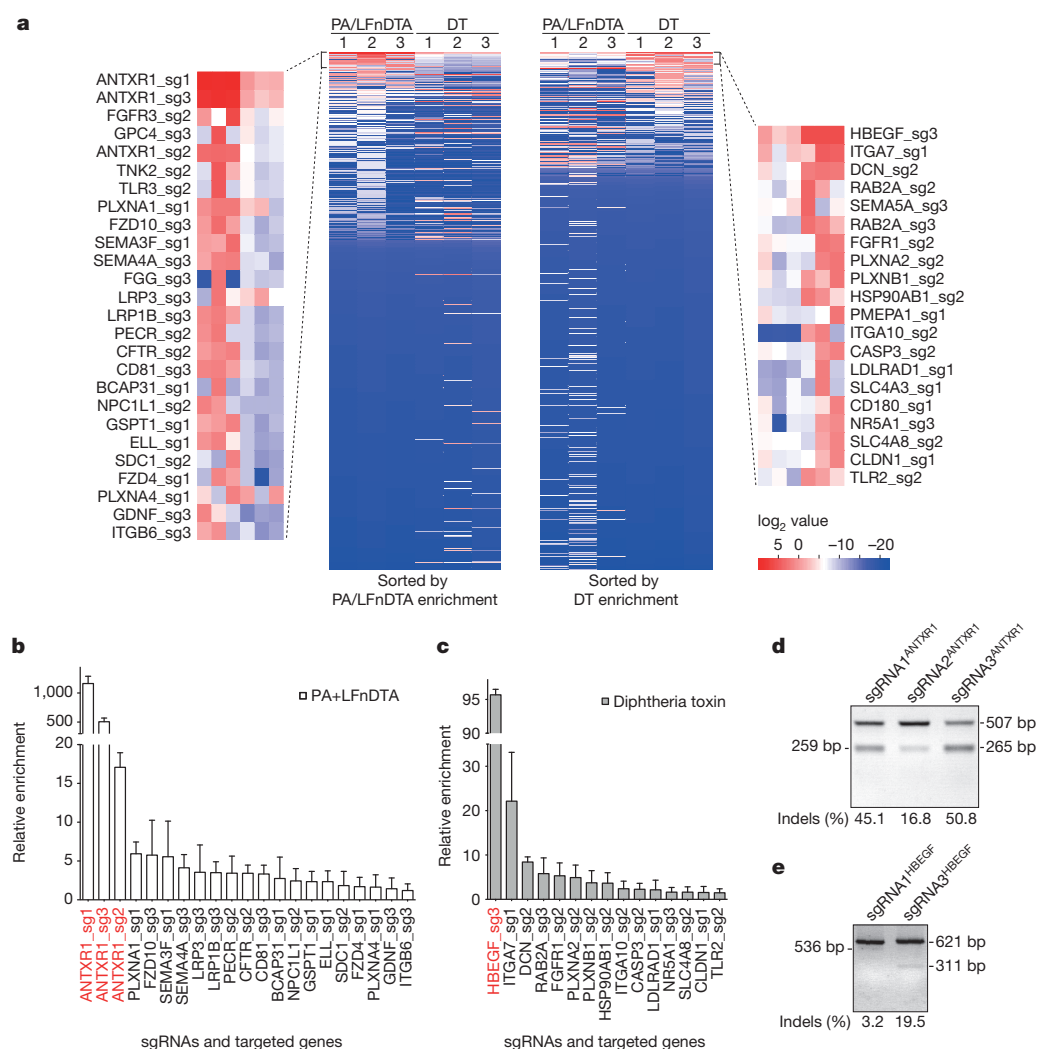


Figure 3 | Screening for essential genes for PA/LFnDTA and diphtheria toxin (DT) toxicity. **a**, Heat maps generated from sgRNA library screening of PA/LFnDTA and diphtheria toxins and high-throughput sequencing analysis. The two heat maps are sorted separately by average enrichment level of sgRNAs obtained from three parallel library screenings with PA/LFnDTA (left) and diphtheria toxin (right). The enrichment level is calculated as $\log_2(\text{Exp}/\text{Ctrl})$ of the normalized read counts. The partial enlarged images show the top sgRNAs with their targeted genes determined from the screens of two toxins. **b**, **c**, Enrichment of candidate sgRNAs and their targeted genes from PA/LFnDTA (**b**) and diphtheria toxin (**c**) screening. sgRNA enrichment was calculated as the average of normalized experimental read counts divided by the control, and was plotted as the mean of three replicates in descending order. sgRNAs targeting positive genes known to be important for anthrax (**b**) or diphtheria toxin (**c**) toxicity are highlighted in red. Data are presented as the mean \pm s.e.m. ($n = 3$). **d**, **e**, T7E1 digestion assays indicating indels induced by indicated sgRNAs targeting ANTXR1 (**d**) or HBEGF (**e**).

multiple *RAB2A*-targeting sgRNAs in HeLa_{OC}-SC cells (Extended Data Fig. 8).

To assess further the candidate genes isolated from our library screening and the above validation procedure, we chose the *PECR* gene for further extensive study. *PECR* gene knockouts were isolated through the CRISPR/Cas9 system using two different sgRNAs, and the complete disruption of *PECR* was confirmed by Sanger sequencing (Extended Data Fig. 9a, b). The complete loss of *PECR* in both independently acquired clones conferred on HeLa cells increased levels of resistance to PA/LFnDTA, but not to diphtheria toxin (Fig. 4d and Extended Data Fig. 9c). *PECR*-specific short hairpin RNAs (shRNAs) also generated a resistant phenotype in HeLa cells to PA-mediated killing (Extended Data Fig. 9d), further demonstrating the positive role of *PECR* in anthrax toxicity.

In addition to *ANTXR1*, we also designed three sgRNAs targeting the other anthrax toxin receptor gene, *ANTXR2* (ref. 19). Only two sgRNAs (sgRNA2 and sgRNA3) were retained in the library. Although sgRNA2^{ANTXR2} induced indels at an efficiency of 21.3% by T7E1 assay, neither sgRNA generated knockout cells that survived the PA/LFnDTA toxin treatment (Supplementary Tables 3 and 4). Using two pairs of *ANTXR1*- and *ANTXR2*-targeting TALENs construct¹⁸, we introduced each pair of plasmids into HeLa and HEK293T cells. From randomly picked clones harbouring either of the two pairs of TALENs constructs, toxin-resistant clones were identified only from the HeLa TALENs^{ANTXR1} (19 of 50) and the HEK293T TALENs^{ANTXR2} (5 of 32), but not from the HeLa TALENs^{ANTXR2} (0 of 50) or the HEK293T TALENs^{ANTXR1} (0 of 32) cells (Extended Data Fig. 10a), suggesting that these two receptors

have different roles in these two cell types. Reverse transcription PCR analysis revealed that the basal expression of *ANTXR2* transcripts was barely detectable in HeLa cells (Extended Data Fig. 10b). This explained why ANTXR1, but not ANTXR2, has such a dominant role in HeLa cells. As a more potent receptor²⁰, ANTXR2 instead of ANTXR1 has a more significant role when the expression of both genes was comparable in HEK293T cells (Extended Data Fig. 10).

Although arrayed and pooled screens using RNA interference (RNAi) libraries have already been developed and are widely used for systematic genetic studies in mammalian cells, they do have certain limitations—in particular, the RNAi-based downregulation of a particular gene is not always sufficient to cause the phenotypic change of interest²¹. Therefore, methods based on gene knockout screening are highly desirable. Here, we established an effective method for the construction of a pooled gene-knockout library and demonstrated that this CRISPR-based strategy can be seamlessly implemented in functional genomics when combined with deep-sequencing analysis. The broad application of this scalable genetic tool may further augment the power of the CRISPR/Cas9 system in studying gene functions in a high-throughput fashion.

While our manuscript was under revision, two parallel studies^{8,9} reported a similar lentiviral-based pooled screening using the CRISPR/Cas9 system, both of which developed large libraries covering over 18,000 (ref. 9) and 7,000 (ref. 8) genes. The genome-scale screening of the CRISPR/Cas9 library is valuable; however, it might be technically challenging to manage the library. Our approach presents an alternative way—that is, via a focused library—to conduct functional genomics,

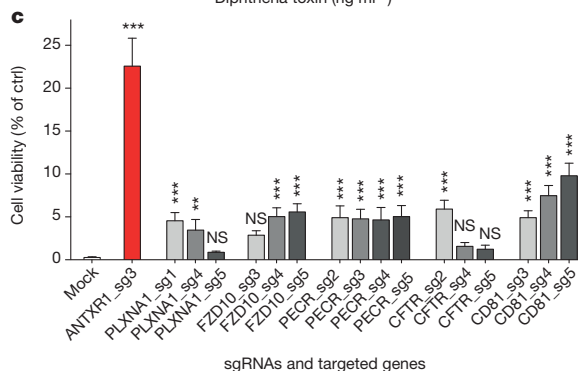


Figure 4 | Genetic validation of candidate genes. **a, b, d,** Effects of *ANTXR1* (**a**), *HBEGF* (**b**) and *PECR* (**d**) deficiency on PA/LFnDTA- and diphtheria-toxin-triggered cytotoxicity in HeLa cells. The cells were treated with different doses of toxins, the XTT cell viability assay and LDH lethality assay were performed 48 h after toxin treatment. Data are presented as the mean \pm s.d. ($n = 3$). **c,** Relative cell viability for HeLa_{OC}-SC cells transfected with indicated sgRNA constructs. Four days after transfection, cells were incubated with

and is particularly advantageous for knowledge-based screening. The other difference is that ref. 9 chose to combine Cas9 and sgRNA expression into a single vector, as it is time-consuming to generate the cell lines that express Cas9. However, we found that individual clones harbouring Cas9 varied in their efficiency to generate sgRNA-mediated indels. It is therefore important to select the best single clone for library construction for certain cell types. In addition, the universal background of Cas9 derived from a single clone is beneficial for the accurate assessment of sgRNA contribution manifested by the final counts from the deep-sequencing analysis.

METHODS SUMMARY

The lentiviral plasmid expressing OCT1 and Cas9 was constructed using Gibson's method^{22,23} in pLenti-CMV-BSD, and the lentiviral backbone used for the sgRNA library cloning was constructed in the pLL3.7 vector (Addgene, Inc.). The oligonucleotides used for the sgRNA library construction were individually designed (Supplementary Table 2) and synthesized (Ruibiotech, Inc.). HeLa_{OC}-SC cells infected by the library viruses were sorted by FACS (MOFLO, Cytomation) based on the presence of EGFP. The library cells were screened by three rounds of toxin treatment: PA (70 ng ml⁻¹) plus LFnDTA (50 ng ml⁻¹); and diphtheria toxin (7.5 ng ml⁻¹). Three replicates were arranged for each toxin screening. The genomic DNA of each replicate as well as the original library was isolated using the DNeasy Blood and Tissue kit (Qiagen). sgRNA-coding regions integrated into the chromosomes were PCR-amplified (TransTaq DNA Polymerase High Fidelity, TransGen) through 26 cycles of reaction with primers annealed to the flanking

PA (70 ng ml⁻¹) plus LFnDTA (50 ng ml⁻¹) for 48 h. Images of toxin-treated and untreated cells were taken (Extended Data Fig. 7) for cell number survey by ImageJ program. The cell viability was calculated as the percentage of surviving cells after toxin treatment. The data are the mean \pm s.d. ($n = 6$); ** $P < 0.01$; *** $P < 0.001$; NS, not significant; Dunnett's multiple comparison test, one-way ANOVA.

sequences of the sgRNAs (Supplementary Table 5) before being subjected to high-throughput sequencing analysis (Illumina HiSeq 2500).

PA and LFnDTA, a surrogate of LF consisting of the amino-terminal domain of LF fused to the catalytic subunit of diphtheria toxin, were produced by using plasmid pET-22b-PA and pET-15b-LFnDTA (Addgene plasmid 11079 and 11075), respectively. The diphtheria toxin was purchased from List Biological Laboratories, Inc. Cytotoxicity assays were performed as described²⁴ with XTT (Roche) according to the product manual. LDH staining and detection were performed as described in the product instruction (CytTox96, Promega).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 December 2013; accepted 24 February 2014.

Published online 9 April 2014.

1. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
2. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnol.* **25**, 778–785 (2007).
3. Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307 (2011).
4. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnol.* **29**, 143–148 (2011).
5. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnol.* **29**, 149–153 (2011).
6. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

7. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
8. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
9. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
10. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
11. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
12. Lin, B. R. & Natarajan, V. Negative regulation of human U6 snRNA promoter by p38 kinase through Oct-1. *Gene* **497**, 200–207 (2012).
13. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
14. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* **39**, 9283–9293 (2011).
15. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
16. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
17. Naglich, J. G., Metherall, J. E., Russell, D. W. & Eidels, L. Expression cloning of a diphtheria toxin receptor: identity with a heparin-binding EGF-like growth factor precursor. *Cell* **69**, 1051–1061 (1992).
18. Yang, J. *et al.* ULtiMATE system for rapid assembly of customized TAL effectors. *PLoS ONE* **8**, e75649 (2013).
19. Scobie, H. M., Rainey, G. J., Bradley, K. A. & Young, J. A. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl Acad. Sci. USA* **100**, 5170–5174 (2003).
20. Liu, S. *et al.* Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin *in vivo*. *Proc. Natl Acad. Sci. USA* **106**, 12424–12429 (2009).
21. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nature Biotechnol.* **29**, 79–83 (2011).
22. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
23. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361 (2011).
24. Wei, W., Lu, Q., Chaudry, G. J., Leppla, S. H. & Cohen, S. N. The LDL receptor-related protein LRP6 mediates internalization and lethality of anthrax toxin. *Cell* **124**, 1141–1154 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Xi for providing the Cas9-encoding construct, F. Tang for technical advice regarding the high-throughput sequencing analysis, Z. Jiang for aid with the cell lines, Y. Yang for assistance in primer design, Y. Wang for data analysis, and D. Trono for aid with the lentivirus packaging system. We thank Peking University High-throughput Sequencing Center operated by BIOPIIC. This work was supported by funds from the National Basic Research Program of China (2010CB911800), the National Science Foundation of China (NSFC31170126, NSFC31070115), and the Peking-Tsinghua Centre for Life Sciences.

Author Contributions W.W., Y.Z., S.Z. and C.C. conceived the methodology and designed the experiments. Y.Z., S.Z., C.C., P.Y. and C.L. performed the experiments. W.W., Y.Z., S.Z., C.C., P.Y. and Y.H. analysed the data. W.W., C.C., Y.Z. and S.Z. wrote the manuscript with help from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W. (wswei@pku.edu.cn).

METHODS

Cells and reagents. HT1080, HeLa and HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS, Invitrogen) with 5% CO₂ in 37 °C.

Plasmid construction. The lentiviral sgRNA vector (pLenti-sgRNA-Lib) was constructed by substituting its original U6 promoter in pLL3.7 (Addgene, Inc.) for the human U6 promoter, *ccdB* cassette and sgRNA scaffold. The gene *OCT1* was PCR-amplified from a human cDNA library (Abclonal, Inc.) and fused with a 2A-linked²⁵ humanized *Cas9* gene²⁶, followed by an IRES (pIRES, Invitrogen)-linked blasticidin-resistance gene. The construction of this plasmid (pLenti-OC-IRES-BSD) was accomplished using Gibson's method^{22,23}, and was verified by sequencing analysis. Details regarding these two constructs are given in Supplementary Fig. 1.

Construction of the CRISPR/Cas9 sgRNA library. We created a library targeting 291 human genes based on our previous shRNA library screening results (our own unpublished data). We have subjectively added additional genes based on previous knowledge and our research interests. We preferentially selected genes encoding cell surface proteins, and those important for endocytosis, trafficking and cell death. The oligonucleotides for each sgRNA-coding sequence for the library were individually designed (Supplementary Table 2) and synthesized (Ruibiotec, Inc.). Paired oligonucleotides were mixed in 96-well plates to the final concentration of 9 μM in 50 μl of TransTaq HiFi Buffer II (1×) for annealing. These annealed oligonucleotide pairs were then mixed, phosphorylated and ligated into the lentiviral sgRNA vector using the Golden Gate method. The ligation mixture was transformed into Trans1-T1 competent cells (Transgen) to obtain the library plasmids.

Lentiviruses of the sgRNA library were obtained by co-transfection of library plasmids with two viral packaging plasmids psPAX2 and pMD2.G (from Didier Trono of EPFL) into HEK293T using the polyethylenimine (PEI) method. HeLa (OCT1-Cas9) cells were infected by the library viruses, followed by FACS (MOFLO, Cytomation) for EGFP-positive cells, 48 h after infection. The viral titre used was 0.05, and the total cell number for viral infection was 4×10^7 .

CRISPR/Cas9 sgRNA library screening. A total of 3×10^7 sgRNA library cells were plated onto 100-mm Petri dishes at 1×10^6 cells per plate. After 24 h, the cells were treated by toxins at an appropriate concentration: PA (70 ng ml⁻¹) plus LFnDTA (50 ng ml⁻¹); and diphtheria toxin (7.5 ng ml⁻¹). For each toxin screening, five plates of cells were grouped together as one replicate, and three replicates were arranged for each toxin screen. For each round of screening, the library cells were incubated with the appropriate toxins for 2 days before being changed to fresh DMEM. The surviving cells were re-seeded for growth and subjected to another round of toxin screening. After three rounds of toxin treatment, the resistant clones and the original untreated library cells were collected separately for genomic DNA extraction, followed by PCR amplification of the sgRNA-coding region and deep-sequencing analysis.

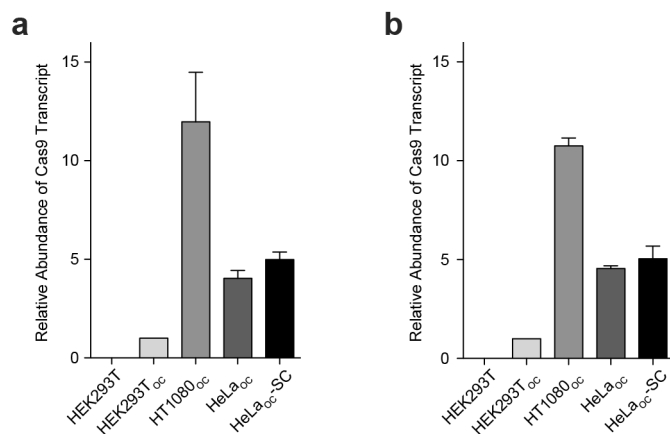
Identification of candidate sgRNA sequences. The genomic DNA of six replicates was isolated from 5×10^6 cells using the DNeasy Blood and Tissue kit (Qiagen). sgRNA-coding regions integrated into the chromosomes were PCR-amplified (TransTaq DNA Polymerase High Fidelity, TransGen) with 26 cycles of reaction using primers annealed to the flanking sequences of the sgRNAs (Supplementary Table 4). In this work, 0.5 μg genomic DNA was used in each 50-μl PCR reaction, 16 PCR reactions were performed for each replicate, and different replicates were indexed with different barcodes. The PCR products (16 tubes) of each replicate were pooled and purified with DNA Clean & Concentrator-5 (Zymo Research Corporation), followed by high-throughput sequencing analysis (Illumina HiSeq 2500).

Statistical analysis. The R software package from Bioconductor, DESeq2, was used to perform a statistical analysis of the sequencing data. The enrichment of sgRNA was ranked by the average fold change of normalized counts: reads^{Exp}/reads^{Ctrl}. The adjusted *P* value was calculated to evaluate the data quality. The criterion for the selection of candidate sgRNAs and their targeted genes is an adjusted *P* ≤ 0.05 for which the fold changes are greater than that of the negative control plus its standard deviation (mean + s.d.). sgRNAs targeting *ANTXR1* and *HBEGF* were chosen to serve as negative controls for the diphtheria toxin and PA/LFnDTA screening, respectively.

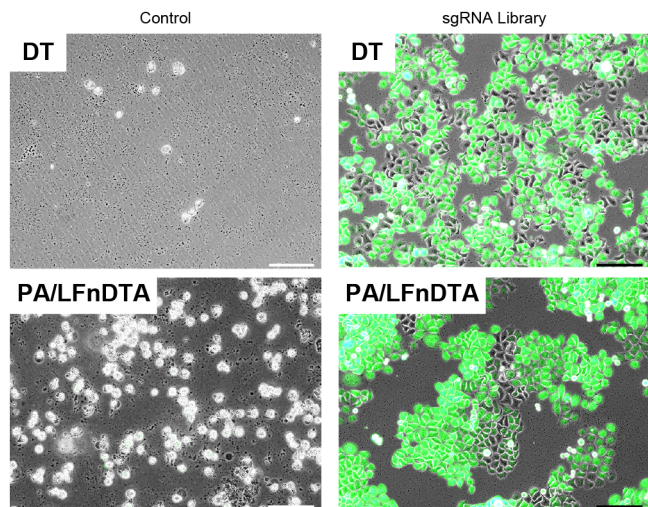
Cytotoxicity assay. PA and LFnDTA, a surrogate of LF consisting of the N-terminal domain of LF fused to the catalytic subunit of diphtheria toxin, were produced using plasmid pET-22b-PA and pET-15b-LFnDTA (Addgene plasmid 11079 and 11075), respectively. Diphtheria toxin was purchased from List Biological Laboratories, Inc. Cytotoxicity assays were performed as described²⁴ with XTT (Roche) according to the product manual. LDH staining and detection were performed as described in the product instruction (CytTox96, Promega). The death signal represented by the amount of LDH release was normalized to the wells based on the maximum LDH activity of the total lysed cells. Each data point and related error bar shown in the figures for the XTT or LDH assays represent the average results from three replicates.

Real-time PCR. RNA of cultured cells was extracted using EasyPure RNA kit (Transgen, ER101-01), and the cDNAs were synthesized using PrimeScript 1st Strand cDNA Synthesis kit (TAKARA, 6110A). Real-time PCR was performed with Platinum SYBR green qPCR SuperMix-UDG (Life Technologies, C11733-038) on Stratagene Mx3005P qPCR system. Two pairs of primers were designed to examine the expression level of each gene. β-actin and GAPDH transcript levels were measured as internal controls.

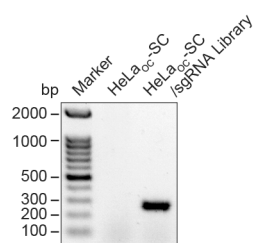
25. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE* **6**, e18556 (2011).
26. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res.* **23**, 465–472 (2013).



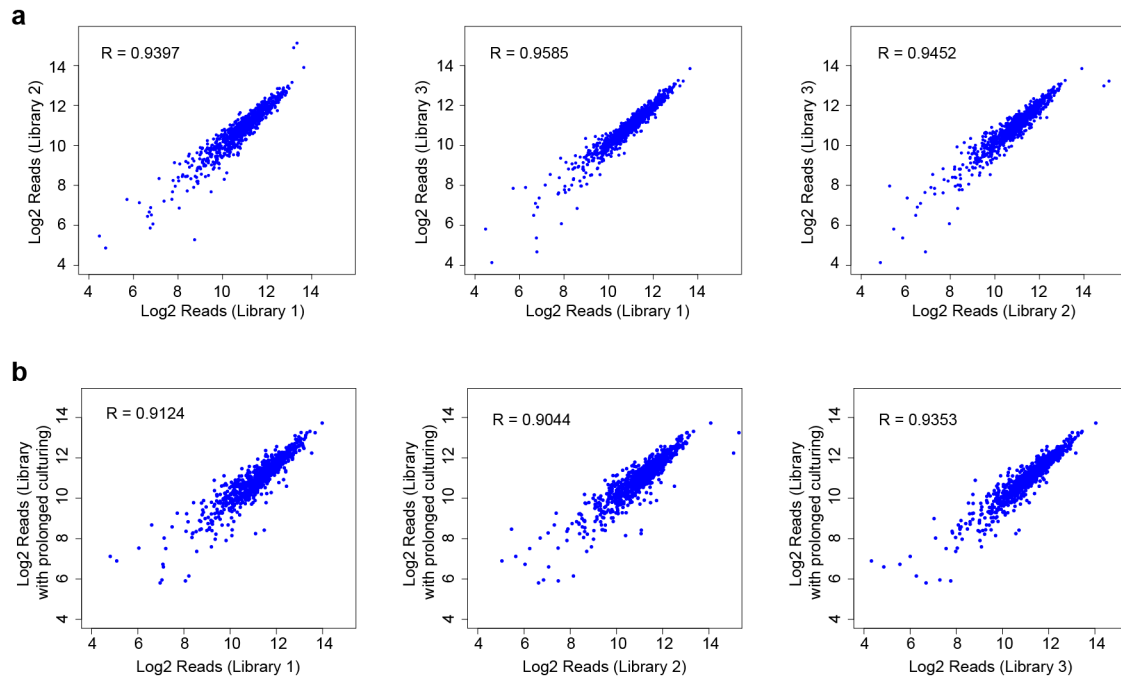
Extended Data Figure 1 | Relative expression level of Cas9 by real-time PCR. Two pairs of primers (**a**, 5'-GGGGGACAGTCTTCACGAGC-3' and 5'-CACGTACATGTCCCTGCCGT-3'; **b**, 5'-GCACGTGGCCCAAATTCTCG-3' and 5'-AGCGGTGGCCTTGCCTATTT-3') were designed to test Cas9 expression level. cDNA of HEK293T serves as the negative control. β -Actin was used as the internal control and the Cas9 expression of HEK293T_{OC} was set as the calibrator. The data are the mean \pm s.d. ($n = 3$).



Extended Data Figure 2 | Images of cells after three rounds of toxin treatment: diphtheria toxin (7.5 ng ml^{-1}) for 60 h or PA/LFnDTA (70 ng ml^{-1} / 50 ng ml^{-1}) for 48 h. The fluorescence signals were superimposed onto white light images. Scale bar: $100 \mu\text{m}$.

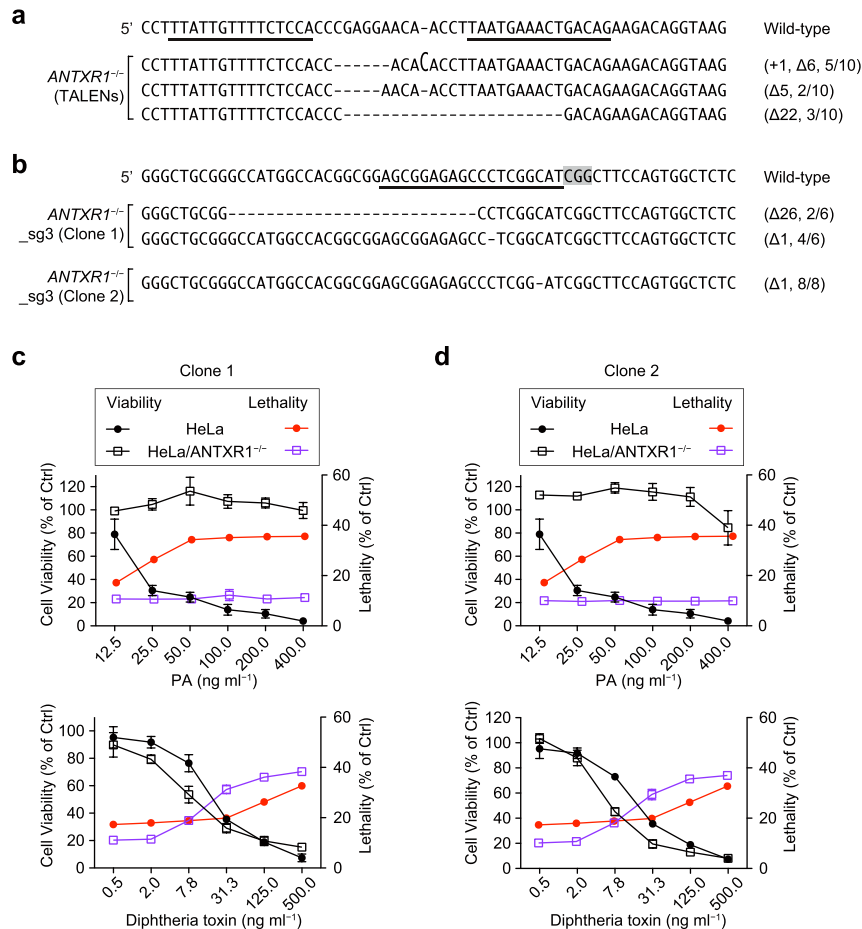


Extended Data Figure 3 | PCR amplification of the sgRNA region from the sgRNA library for deep-sequencing analysis, as indicated by electrophoresis. The genomic DNAs isolated from the indicated cells were used as templates for PCR amplification.



Extended Data Figure 4 | Scatter plots of \log_2 normalized read counts between two sgRNA libraries. **a**, Scatter plots of \log_2 normalized read counts between two of the three replicated sgRNA libraries. **b**, Scatter plots of \log_2

normalized read counts between two sgRNA libraries with or without prolonged culturing (14 days), and three replicates of the original library were used for the comparison.



Extended Data Figure 5 | Effect of *ANTXR1* deficiency on PA/LFnDTA- and diphtheria-toxin-triggered cytotoxicity in HeLa cells. **a**, Partial sequences of *ANTXR1* gene in genome containing TALENs binding regions (underlined) and the mutated alleles from 10 randomly selected TALENs clones (in HeLa cells). The tall letter indicates insertion, and the dashes indicate deletions in this and other figures. **b**, DNA sequencing analysis of mutated alleles in the human *ANTXR1* locus. Partial coding sequences of targeted genes in the genome containing the CRISPR/Cas9 sgRNA3^{ANTXR1} binding region (underlined) and sequence analysis of the mutated alleles from two

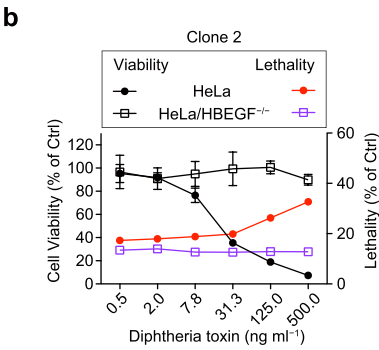
independent clones are shown. Primers used for the PCR reactions are listed in Supplementary Table 5, and the shaded nucleotides represent the PAM sequences that guide the Cas9 for DNA recognition and cleavage in this and other figures. **c**, **d**, Susceptibility assay of two HeLa *ANTXR1*^{-/-} clones on PA/LFnDTA- and diphtheria-toxin-triggered cytotoxicity. The cells were treated with different doses of toxins, and the XTT cell viability assay and LDH lethality assay were performed 48 h after toxin treatment. Data are presented as the mean \pm s.d. ($n = 3$).

a

5' GCAGTTCTCTCGGCACTGGT-GACTGGCGAGAGCCTGGAGCGGCTTCGGAG Wild-type

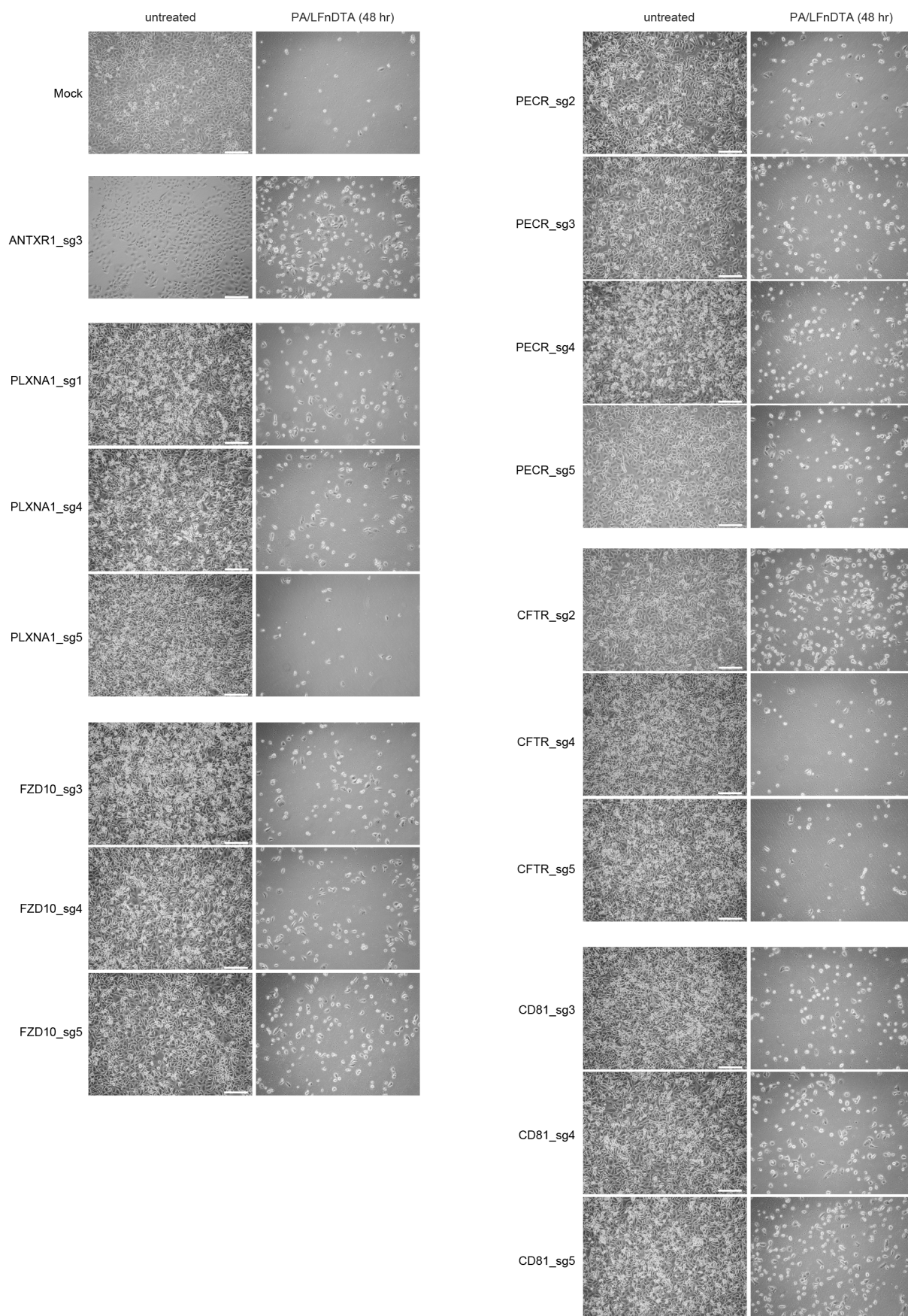
HBEGF^{-/-} _sg3 (Clone 1) [GCAGTTCTCTCGGCACTGG-----CGAGAGCCTGGAGCGGCTTCGGAG (Δ7, 2/7)
GCAGTTCTCTCGGCACTGGT GACTGGCGAGAGCCTGGAGCGGCTTCGGAG (+1, 5/7)

HBEGF^{-/-} _sg3 (Clone 2) [GCAGTTCTCTCGGCACTGGT GACTGGCGAGAGCCTGGAGCGGCTTCGGAG (+1, 7/7)



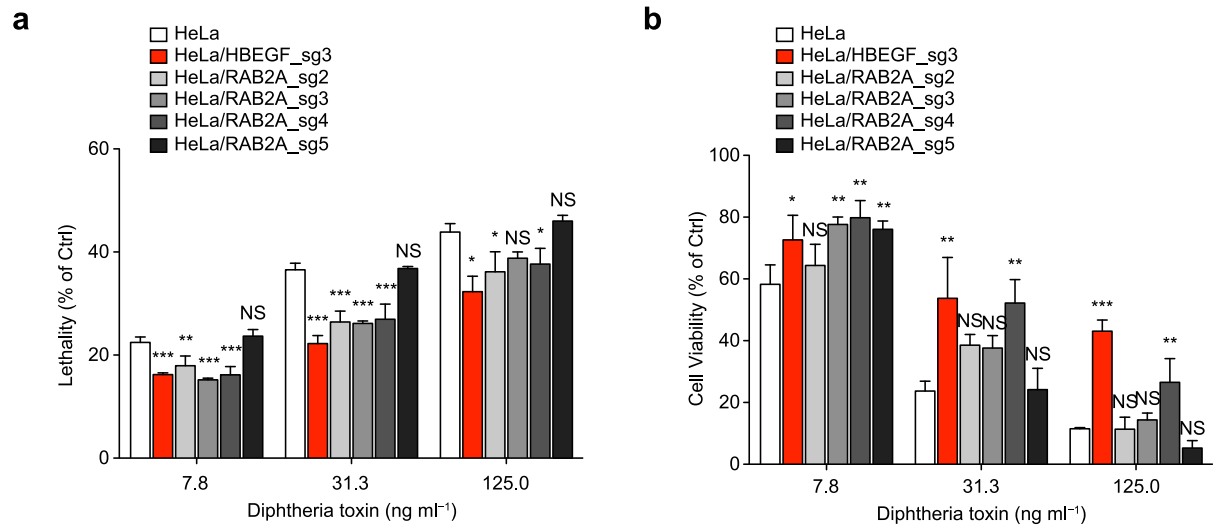
Extended Data Figure 6 | Effect of *HBEGF* deficiency on diphtheria-toxin-triggered cytotoxicity in HeLa cells. **a**, DNA sequencing analysis of mutated alleles in the human *HBEGF* locus. Partial coding sequences of targeted genes in the genome containing the CRISPR/Cas9 sgRNA3^{*HBEGF*} binding region (underlined) and sequencing analysis of the mutated alleles from two

independent clones. **b**, Susceptibility assay of HeLa *HBEGF*^{-/-} clone 2 on diphtheria-toxin-triggered cytotoxicity. The cells were treated with different doses of toxins, and the XTT cell viability assay and LDH lethality assay were performed 48 h after toxin treatment. Data are presented as the mean ± s.d. ($n = 3$).



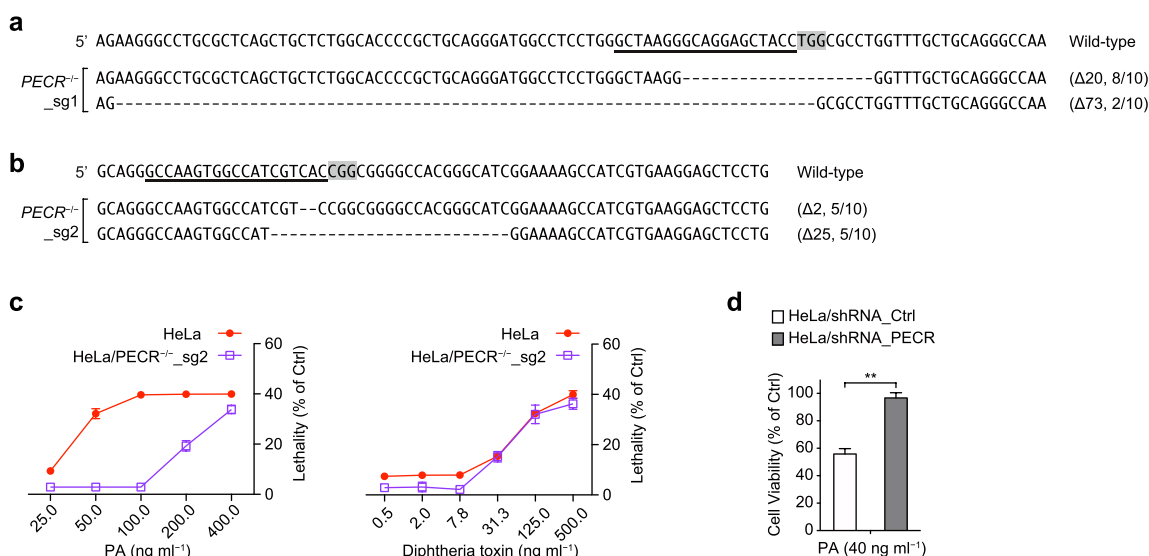
Extended Data Figure 7 | Light microscopic images of HeLa_{OC}-SC cells transfected with indicated sgRNA constructs. Four days after transfection, cells were incubated with PA (70 ng ml⁻¹) plus LFnDTA (50 ng ml⁻¹) for 48 h.

The representative images for every condition (six images for each) are shown. Scale bar, 200 µm.



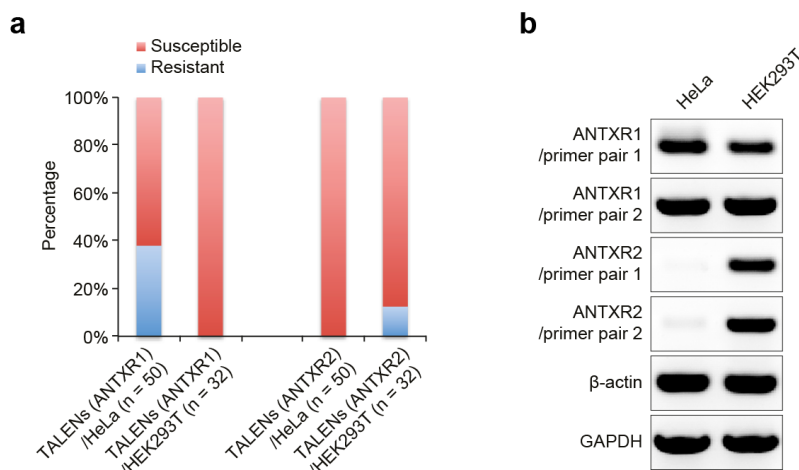
Extended Data Figure 8 | Effects of *RAB2A* sgRNAs on diphtheria-toxin-triggered cytotoxicity in HeLa cells. HeLa_{OC-SC} cells were transfected with indicated sgRNA constructs. Four days after transfection, the LDH lethality assay (a) and the XTT cell viability assay (b) were performed 48 h after toxin

treatment. HBEGF_sg3, labelled in red, was used as the positive control. The data are the mean \pm s.d. ($n = 3$); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant; Dunnett's multiple comparison test, one-way ANOVA.



Extended Data Figure 9 | Effect of *PECR* deficiency on PA/LFnDTA- and diphtheria-toxin-triggered cytotoxicity in HeLa cells. **a, b**, DNA sequencing analysis of mutated alleles in the human *PECR* locus. Partial coding sequences of targeted genes in the genome containing the CRISPR/Cas9 sgRNA1^{*PECR*} (**a**) and sgRNA2^{*PECR*} (**b**) binding region (underlined) and sequencing analysis of the mutated alleles from two independent clones are shown. **c**, Susceptibility assay of HeLa *PECR*^{-/-}sg2 clone on PA/LFnDTA- and diphtheria-toxin-triggered cytotoxicity. The cells were treated with different doses of toxin, and the LDH lethality assay was performed 48 h after toxin treatment. Data are presented as the mean \pm s.d. ($n = 3$). **d**, Effect of *PECR*-targeting shRNA smart pool on HeLa cells to PA/LFnDTA toxicity. Three different shRNAs (TRCN0000046538: 5'-CCGGGCAGTGGGAATACGGATCAATTCTC

GAGAATTGATCCGTATTCCACTGCTTTTTG-3', TRCN0000046539: 5'-CCGGCCTCACCAATCTTTAGCTTTCTCGAGAAAGCTAAA GATTTGGTGAGGTTTTTG-3', and TRCN0000046541: 5'-CCGGCCCGCTAAACGAATTGGTGTTCTCGAGAACACCAATTC GTTTAGCGGGTTTTTG-3') from MISSION shRNA library (Sigma, Inc.) were chosen, and the MISSION shRNA control vector (SHC002) was used, which contains a sequence that engages with RISC without targeting any known mammalian gene. After transient transfection, 1.5 $\mu\text{g ml}^{-1}$ puromycin was added to obtain the stable shRNA expression cells. The cells were treated with PA (40 ng ml^{-1}) plus LFnDTA (50 ng ml^{-1}), and the XTT cell viability assay was performed 48 h after toxin treatment. Data are presented as the mean \pm s.d., $n = 3$, $**P < 0.01$; t -test.



Extended Data Figure 10 | Effects of *ANTXR1* and *ANTXR2* gene expression on the susceptibility of HeLa and HEK293T cells to PA/LFnDTA. **a**, Effects of *ANTXR1* and *ANTXR2* gene knockout on the susceptibility of the indicated cells to the treatment of PA and LFnDTA. Details of the TALENs constructs used for gene targeting have been previously described¹⁶. Cell clones transfected with TALENs were randomly picked and cultured with PA (100 ng ml⁻¹) and LFnDTA (50 ng ml⁻¹) for 3 days to determine their sensitivity to toxins. **b**, Level of *ANTXR1* and *ANTXR2* transcripts in HeLa and HEK293T cells was measured by RT-PCR, and β-actin and GAPDH were included as internal controls. To ensure accuracy, two pairs of primers were used to measure *ANTXR1* and *ANTXR2* transcript levels,

respectively. The primers used in PCR reactions were: for *ANTXR1*, primer pair 1 5'-GGAAACGGCTTCCGACATGC-3' and 5'-TGTTGTGGGGCCTGATG CAA-3', primer pair 2 5'-GCCCCGATTGCGGACAGTAA-3' and 5'-GCTGACCTGGAGTGCAGCTT-3'; for *ANTXR2*, primer pair 1 5'-TTGGACGGTCTGGTGCCATC-3' and 5'-TCCTCCCCACACA GAACT-3', and primer pair 2 5'-ATGTGGTGGTTTTGGCCCCCT-3' and 5'-GTGTGGGTTTGGGTGCGAGGT-3'; for β-actin, 5'-AGAGCT ACGAGCTGCCTGAC-3' and 5'-AGCACTGTGTTGGCGTACAG-3'; for GAPDH, 5'-ACCACAGTCCATGCCATCAC-3' and 5'-TCCACCACCC TGTGCTGTA-3'.

Copper is required for oncogenic BRAF signalling and tumorigenesis

Donita C. Brady¹, Matthew S. Crowe¹, Michelle L. Turski¹, G. Aaron Hobbs², Xiaojie Yao³, Apirat Chaikwad⁴, Stefan Knapp⁴, Kunhong Xiao³, Sharon L. Campbell², Dennis J. Thiele¹ & Christopher M. Counter^{1,5}

The BRAF kinase is mutated, typically Val 600→Glu (V600E), to induce an active oncogenic state in a large fraction of melanomas, thyroid cancers, hairy cell leukaemias and, to a smaller extent, a wide spectrum of other cancers^{1,2}. BRAF^{V600E} phosphorylates and activates the MEK1 and MEK2 kinases, which in turn phosphorylate and activate the ERK1 and ERK2 kinases, stimulating the mitogen-activated protein kinase (MAPK) pathway to promote cancer³. Targeting MEK1/2 is proving to be an important therapeutic strategy, given that a MEK1/2 inhibitor provides a survival advantage in metastatic melanoma⁴, an effect that is increased when administered together with a BRAF^{V600E} inhibitor⁵. We previously found that copper (Cu) influx enhances MEK1 phosphorylation of ERK1/2 through a Cu–MEK1 interaction⁶. Here we show decreasing the levels of CTR1 (Cu transporter 1), or mutations in MEK1 that disrupt Cu binding, decreased BRAF^{V600E}-driven signalling and tumorigenesis in mice and human cell settings. Conversely, a MEK1–MEK5 chimera that phosphorylated ERK1/2 independently of Cu or an active ERK2 restored the tumour growth of murine cells lacking *Ctr1*. Cu chelators used in the treatment of Wilson disease⁷ decreased tumour growth of human or murine cells transformed by BRAF^{V600E} or engineered to be resistant to BRAF inhibition. Taken together, these results suggest that Cu-chelation therapy could be repurposed to treat cancers containing the BRAF^{V600E} mutation.

Decreasing *Ctr1* expression suppresses MAPK phenotypes in *Drosophila* and signalling in mammalian cells. In a cupric setting MEK1 also binds Cu, and Cu chelation decreases MEK1/2 kinase activity⁶. Cu and, to a smaller extent, silver—which is isoelectric to cuprous Cu—increased the phosphorylation of ERK2 by MEK1, whereas other tested metals had no effect (Extended Data Fig. 1a). Given these results and the dependence of BRAF mutation-positive cancers on MEK1/2 for tumor growth⁸, we investigated whether decreasing the influx of Cu affects BRAF^{V600E} oncogenesis. BRAF^{V600E} was expressed in *Ctr1*^{+/+} and *Ctr1*^{-/-} mouse embryonic fibroblasts (MEFs)⁹, and intracellular Cu deficiency of the latter was confirmed by increased copper chaperone for superoxide dismutase (CCS) levels¹⁰ (Fig. 1a, b). The BRAF^{V600E}-transformed *Ctr1*^{-/-} MEFs showed decreased phosphorylated ERK1/2 (P-ERK1/2), cell growth, and tumour kinetics, effects that were rescued by expressing CTR1 but not the transport-defective mutant¹¹ CTR1^{M154A} (Fig. 1a–d and Extended Data Figs 2a–c and 10a, b). Thus, the Cu-transport function of CTR1 is required for robust BRAF^{V600E} signalling and tumorigenesis.

To assess whether decreasing the binding of Cu to MEK1 affects BRAF^{V600E}-driven tumorigenesis, targeted mutagenesis was undertaken, which revealed that M187A, H188A, M230A, H239A and one other mutation decreased the ability of MEK1 to bind a Cu-charged resin and phosphorylate ERK1/2 (Extended Data Fig. 3a–c). A metal-catalysed oxidation reaction followed by mass spectrometry identified oxidation at H188, M230A, H239 and two other sites (Fig. 1e and Extended Data Fig. 4), suggesting that these residues reside within 10 Å of a Cu atom¹².

We thus focused on H188, M230 and H239 and included M187 as well, even though the oxidation status of this site could not be determined, it lies adjacent to H188 and was similarly required for Cu binding and kinase activity (Fig. 1e and Extended Data Figs 3c and 4e). These four residues are conserved in MEK2 (Extended Data Fig. 5), which, like MEK1, also bound a Cu-charged resin and was inhibited by tetrathiomolybdate (TTM), a Cu chelator (Extended Data Fig. 1b, c). In the three-dimensional MEK1 structure¹³, these four amino acids also clustered such that each was no more than 12.5 Å from the next (Fig. 1e). Combined mutations at these sites progressively decreased MEK1 kinase activity and affinity for a Cu-charged resin, with mutations at all four sites (copper-binding mutant; CBM) having the largest defect in kinase activity, even in the presence of a constitutively active mutation (S218D/S222D; DD)¹⁴ or excess Cu (Fig. 1f–j). Nevertheless, MEK1^{CBM} was still phosphorylated in BRAF^{V600E}-transformed cells, underwent cooperative unfolding and possessed similar thermostability to that of MEK1, as assessed by both circular dichroism and differential scanning fluorimetry (Fig. 1k, l and Extended Data Figs 6a–d and 10c). To exclude the possibility that the CBM mutation decreased kinase activity independently of Cu binding, we took advantage of the fact that MEK5 is highly similar to MEK1/2, yet lacks two sites that are important for binding Cu (Extended Data Fig. 5) and did not bind the Cu-charged resin or show changes in kinase activity in the presence of Cu or TTM (Fig. 1m, n). Specifically, introducing a CBM-equivalent (CBM-E) mutation into MEK5 was found to have no overt effect on the ability of MEK5 to phosphorylate substrates myelin basic protein (MBP) or ERK5 *in vitro* or in cells (Fig. 1o, p). Given these results, we tested and found that the decrease in P-ERK1/2 and tumour growth of BRAF^{V600E}-transformed *Ctr1*^{+/+} MEFs due to knockdown of endogenous *Mek1* messenger RNA by short hairpin RNA (shRNA) (Fig. 1k) was rescued by expressing wild-type but not CBM mutant MEK1 resistant to shRNA-mediated knockdown (Fig. 1l, q and Extended Data Fig. 10c). Thus, under normal Cu homeostasis, inhibiting Cu binding in MEK1 retards BRAF^{V600E}-driven signalling and tumorigenesis.

We next tested whether bypassing the requirement of MEK1/2 for Cu restores BRAF^{V600E}-driven signalling and tumorigenesis to *Ctr1*^{-/-} MEFs. MEK1 was engineered to be independent of Cu by fusing the ERK1/2-binding region of MEK1 protein¹⁵ to the kinase domain of MEK5 protein, rendered active (S311D/T315D; DD)¹⁶ because MEK5 is a substrate of MEKK2 and MEKK3 (ref. 17). Expressing MEK1–MEK5^{DD} resulted in high levels of P-ERK1/2 in *Ctr1*^{-/-} MEFs (Fig. 1r, s), in contrast to similarly activated MEK1^{DD}, which remained sensitive to TTM (Fig. 1j). Furthermore, MEK1–MEK5^{DD} restored tumour growth to these *Ctr1*^{-/-} MEFs (Fig. 1t and Extended Data Fig. 10d). Similarly, expressing ERK2^{R67S}, which increases ERK2 autophosphorylation in a MEK1/2-independent fashion¹⁸, ERK2^{D321N}, which renders the kinase insensitive to phosphatases¹⁹, or ERK2^{GOF}, which combines these two mutations, also led to robust ERK1/2 phosphorylation and restored tumour

¹Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. ²Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁴Nuffield Department of Clinical Medicine, Target Discovery Institute and Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, UK. ⁵Department of Radiation Oncology, Duke University Medical Center, Durham, North Carolina 27710, USA.

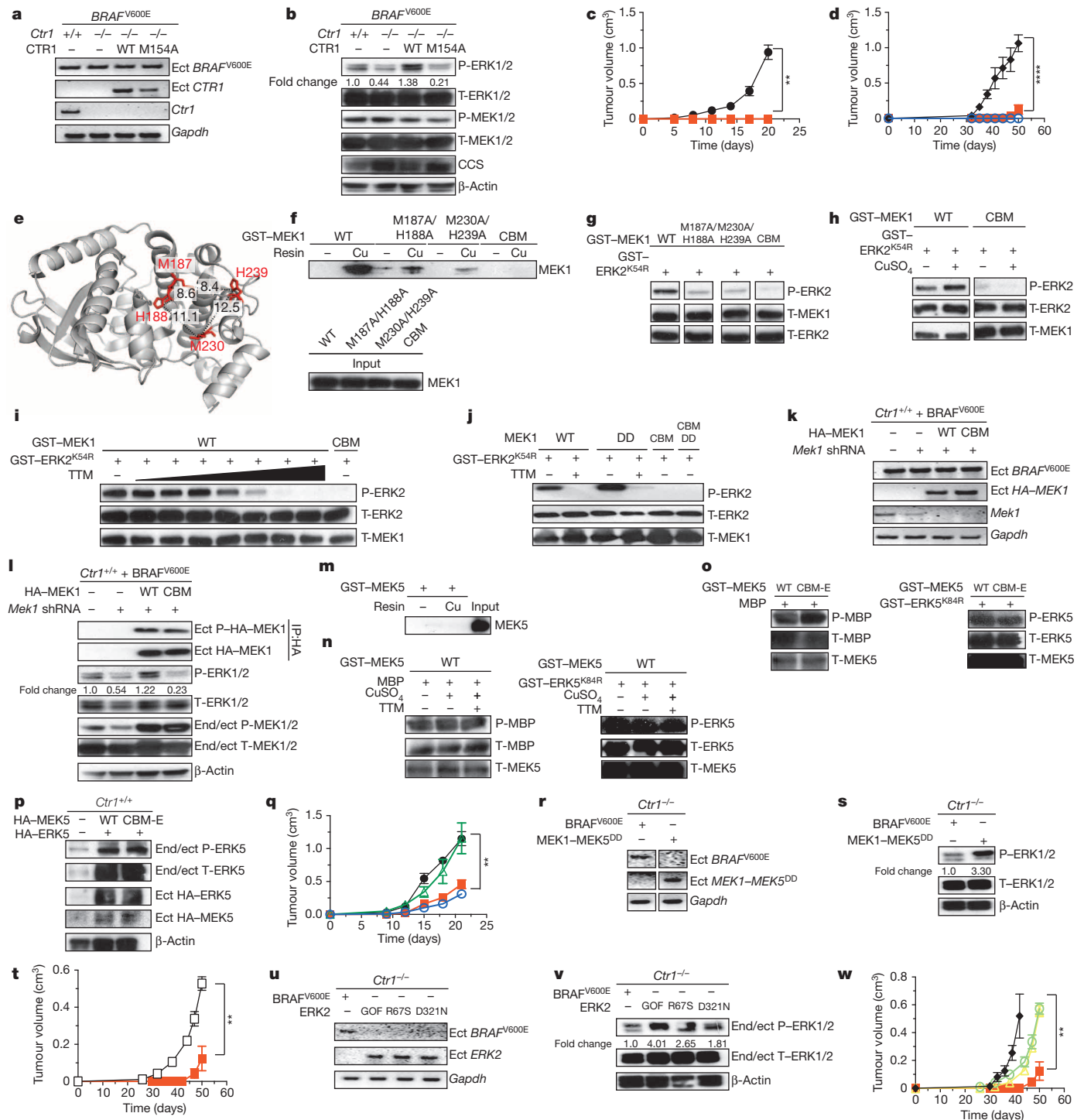


Figure 1 | Binding of Cu to MEK1 promotes MAPK signalling and tumorigenesis by oncogenic BRAF. RT-PCR (a, k, r, u) and immunoblot (b, l, p, s, v) detection of the indicated endogenous, ectopic (ect) or both (end/ect) mRNA and phosphorylated (P) and/or total (T) proteins from cells. IP, immunoprecipitated; WT, wild type. c, d, q, t, w, Mean tumour volume (\pm s.e.m.) versus time of mice injected with BRAF^{V600E}-transformed *Ctrl^{+/+}* (black circles) or *Ctrl^{-/-}* (orange squares) MEFs ($n = 4$) (c), BRAF^{V600E}-transformed *Ctrl^{-/-}* MEFs expressing no transgene (orange squares), CTR1 (black diamonds) or CTR1^{M154A} (blue open circles) ($n = 3$) (d), BRAF^{V600E}-transformed *Ctrl^{+/+}* MEFs expressing scramble shRNA (black circles), *Mek1* shRNA alone (orange squares) or with RNA-mediated-interference-resistant MEK1 (green open triangles) or MEK1^{CBM} (blue open circles) ($n = 3$) (q), *Ctrl^{-/-}* MEFs expressing BRAF^{V600E} (orange squares, $n = 3$) or

MEK1-MEK5^{DD} (black open squares, $n = 4$) (t), or *Ctrl^{-/-}* MEFs expressing BRAF^{V600E} (orange squares, $n = 3$), ERK2^{GOF} (black diamonds, $n = 3$), ERK2^{R67S} (yellow open triangles, $n = 4$) or ERK2^{D321N} (green open triangles, $n = 4$) (w). Results were compared using a one-tailed unpaired *t*-test (c, t) or a one-way ANOVA followed by a Tukey's multi-comparison post test (d, q, w). Two asterisks, $P < 0.01$; four asterisks, $P < 0.0001$. e, MEK1 structure (PDB 3EQD) denoting amino acids M187, H188, M230 and H239 and the intervening space (Å)¹³. f-j, Immunoblot detection of the indicated recombinant proteins bound to a resin charged with or without Cu (f, m) or phosphorylated (P) or total (T) recombinant proteins with or without 50 μ M TTM, a seven-step increase in TTM from 0 to 50 μ M, or either 2.5 molar equivalents or 2.5 μ M CuSO₄ (g, h, i, j, n, o). Gel images are representative of at least two technical replicates.

growth to the *Ctr1*^{-/-} MEFs (Fig. 1u–w and Extended Data Fig. 10d). Thus, while decreasing Cu levels will certainly have pleiotropic effects²⁰ that could affect the MAPK pathway²¹, activating this pathway in a manner independent of Cu nevertheless restores signalling and tumorigenesis in cells deficient in Cu influx.

To investigate the relationship between Cu and endogenous oncogenic BRAF in tumorigenesis, *CTRI* was stably knocked down by shRNA in *BRAF* mutation-positive and mutation-negative melanoma cell lines. Knockdown of *CTRI* decreased P-ERK1/2 levels in all tested cell lines,

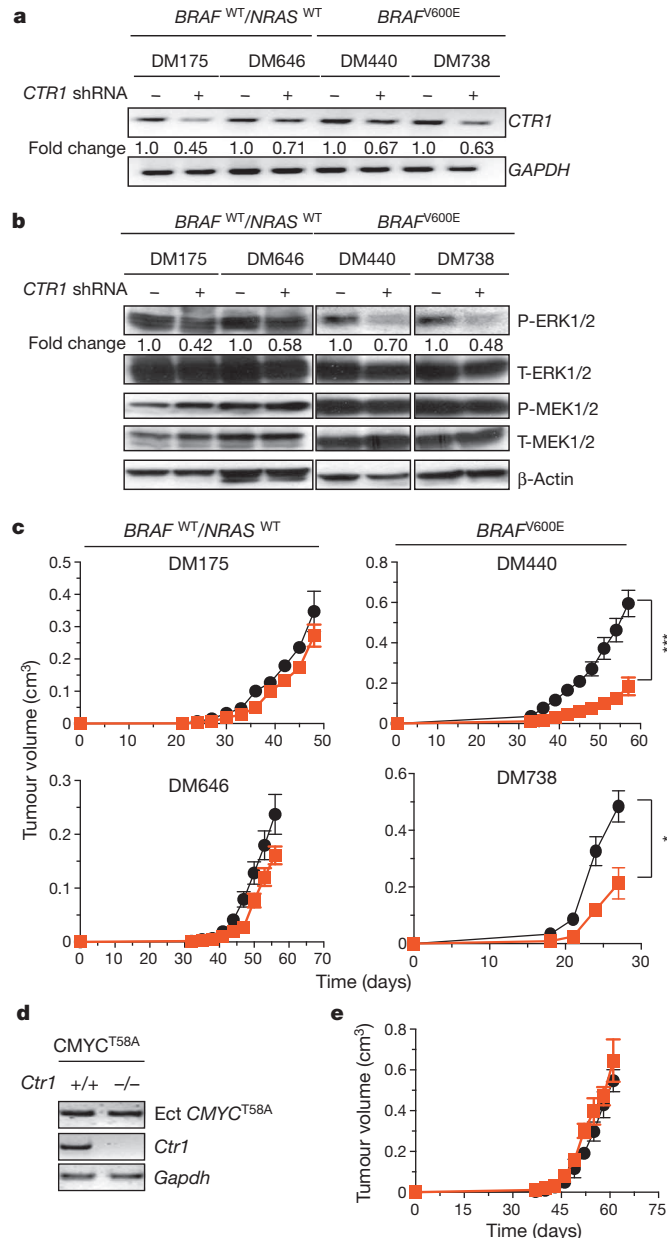


Figure 2 | Knockdown of *CTRI* decreases MAPK signalling and tumorigenesis specifically by oncogenic BRAF. **a–c**, RT–PCR detection of the indicated mRNAs (**a**), immunoblot detection of the indicated phosphorylated (P) or total (T) proteins from cells (**b**), and mean tumour volume (\pm s.e.m.) versus time (**c**) of mice injected ($n = 3$ for DM738; $n = 4$ for others) with the indicated cell lines expressing scramble (black circles) or *CTRI* (orange squares) shRNA. Results were compared using a one-tailed unpaired *t*-test. Asterisk, $P < 0.05$; three asterisks, $P < 0.001$. **d, e**, RT–PCR detection of the indicated endogenous or ectopic (ect) mRNA from cells (**d**) and mean tumour volume (\pm s.e.m.) versus time (**e**) for mice ($n = 3$) injected with CMYC^{T58A}-transformed *Ctr1*^{+/+} (black circles) or *Ctr1*^{-/-} (orange squares) MEFs. Gel images are representative of two technical replicates.

although only the *BRAF* mutation-positive cell lines showed decreased tumour growth (Fig. 2a–c and Extended Data Fig. 10b), consistent with a dependence of BRAF^{V600E} tumorigenicity on the MEK1/2 proteins⁸. Knockdown of *CTRI* also inhibited the tumour growth of one of two tested *NRAS* mutation-positive melanoma cell lines (Extended Data Fig. 7a, b). Finally, loss of *Ctr1* had no effect on the ability of a different oncogene, CMYC^{T58A}, to promote tumour growth (Fig. 2d, e). These results suggest a specific requirement for Cu in cancers that are particularly dependent on the MAPK pathway.

We next evaluated the requirement for Cu influx *in vivo* during spontaneous cancer development. Mice harbouring floxed conditional null or wild-type *Ctr1* alleles²² were crossed into a background having one Cre-activated (CA) conditional oncogenic allele of *Braf* and two floxed conditional null alleles of *Trp53* (*Braf*^{CA/+}; *Trp53*^{flx/flx}, BP)²³, which on intranasal administration of adenovirus expressing Cre recombinase (AdCre) converts *Braf*^{CA} to an oncogenic *Braf*^{V600E} allele and *Trp53*^{flx/flx} to null alleles, leading to the development of lung adenocarcinomas²³. AdCre was administered to cohorts of *Ctr1*^{+/+} and *Ctr1*^{flx/flx} BP mice to induce cancer development and, in the latter case, to convert *Ctr1*^{flx/flx} to null alleles. Tumours arising in both cohorts showed the appropriate recombination of the *Braf*^{CA}, *Trp53*^{flx} and

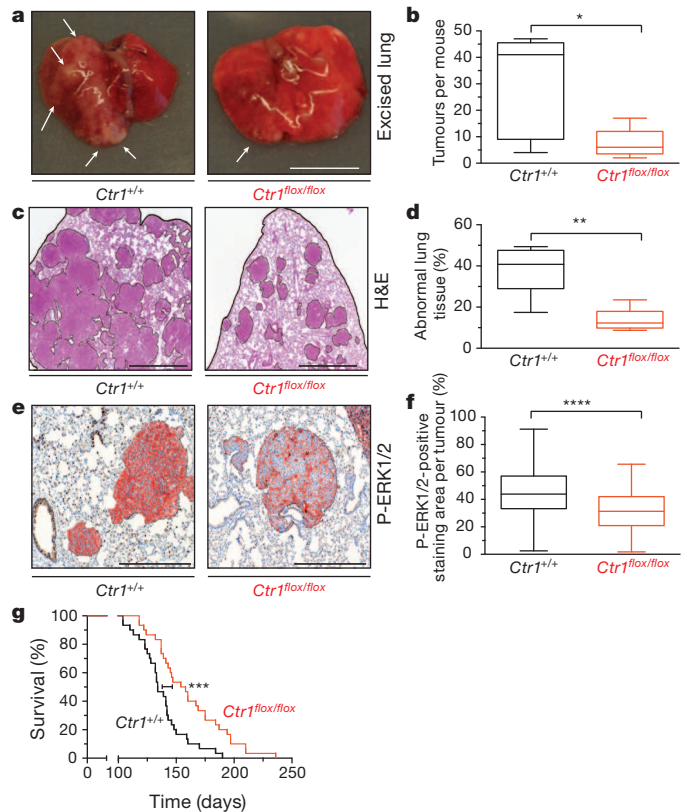
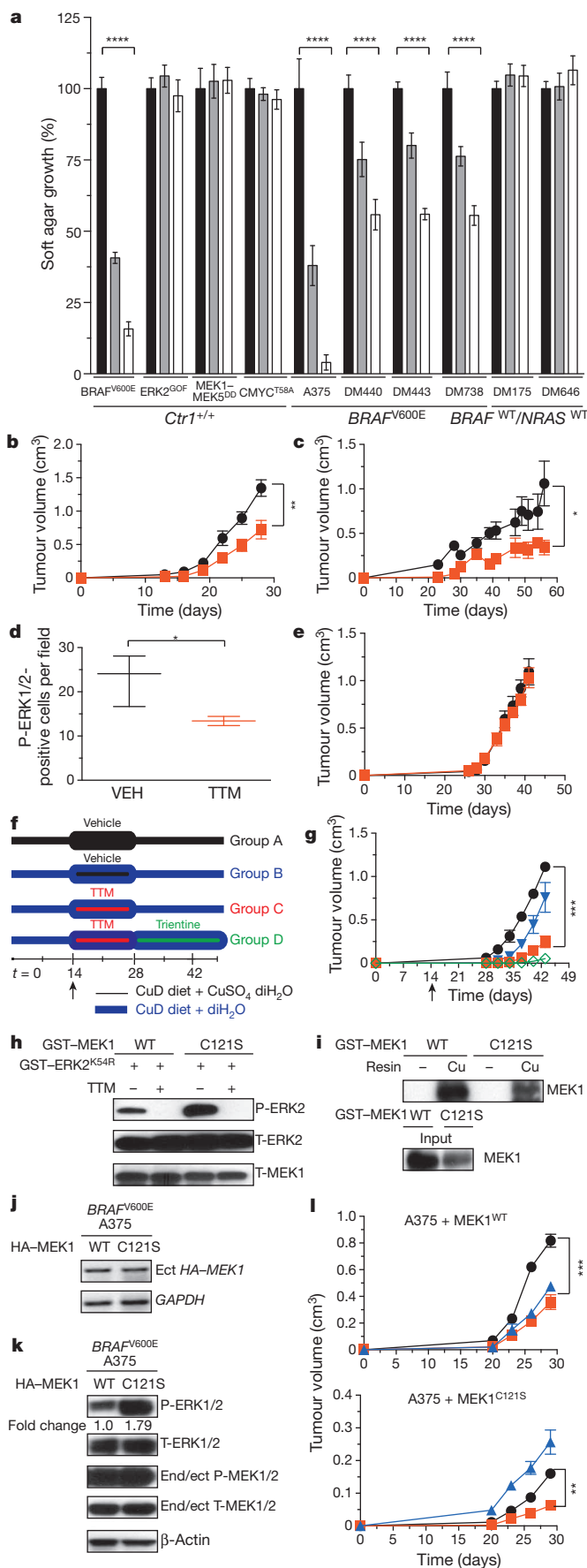


Figure 3 | Genetic ablation of *Ctr1* decreases MAPK signalling and tumorigenesis and extends the lifespan in a mouse model of *BRAF*^{V600E}-driven lung cancer. **a, c, e**, Representative resected (arrows indicate visible lesions; scale bar, 1 cm) (**a**), haematoxylin/eosin-stained (scale bar, 1 mm) (**c**) and immunohistochemical detection of P-ERK1/2 (orange, positive pixels; scale bar, 500 μ m) (**e**) in lungs of *Ctr1*^{+/+} or *Ctr1*^{flx/flx} BP mice (**a, c**, fixed endpoint; **e**, moribundity endpoint). **b, d, f**, Box-and-whiskers plot of number of tumours per mouse ($n = 5$ lungs) (**b**), mean percentage of area of abnormal lung tissue ($n = 5$ lungs) (**d**) and mean percentage of P-ERK1/2 positive-staining area per lung tumour (**f**) of *Ctr1*^{+/+} ($n = 199$ tumours) and *Ctr1*^{flx/flx} ($n = 142$ tumours) BP mice (**b, d**, fixed endpoint; **f**, moribundity endpoint). **g**, Kaplan–Meier survival analysis of *Ctr1*^{+/+} (black line, $n = 30$) and *Ctr1*^{flx/flx} (orange line, $n = 30$) BP mice. Results were compared using a one-tailed unpaired *t*-test (**b, d, f**) or a Mantel–Cox test (**g**). Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$; four asterisks, $P < 0.0001$.



Ctr1^{flax} alleles, and targeted loss of *Ctr1* did not result in a weight loss (Extended Data Fig. 8a, b). *Ctr1*^{flax/flax} BP mice showed a decrease in the number of visible surface lesions, the area of abnormal lung tissue, and P-ERK1/2 staining in tumours and also a 15% survival advantage over *Ctr1*^{+/+} BP mice (Fig. 3a–g and Extended Data Fig. 10a, b). Thus, the loss of *Ctr1* retards BRAF^{V600E}-driven tumorigenesis, leading to a survival advantage.

The Cu chelators D-penicillamine, trientine and TTM are efficacious, long-term treatments for lowering systemic Cu levels in individuals with the Cu-overload disorder Wilson disease⁷. Capitalizing on these drugs, we found that TTM decreased the growth of BRAF^{V600E}, but not ERK2^{GOF}, MEK1–MEK5^{DD} or MYC^{T58A}-transformed MEFs in soft agar. The growth of BRAF mutation-positive melanoma cell lines in soft agar was also inhibited by TTM to various degrees in a dose-dependent manner, whereas BRAF/NRAS mutation-negative lines were resistant to TTM (Fig. 4a and Extended Data Fig. 10e). Oral TTM treatments for only two weeks also decreased the size of tumours in mice injected with either BRAF^{V600E}-transformed MEFs or the BRAF mutation-positive melanoma cell line DM440, with no negative effect on mouse weight (Fig. 4b, c and Extended Data Figs 9 and 10e). This effect could be ascribed to a decrease in MEK1/2 kinase activity, because TTM halved the number of P-ERK1/2-positive cells in BRAF^{V600E}-driven tumours and failed to inhibit tumour growth of ERK2^{GOF}-transformed cells (Fig. 4d, e). To investigate more aggressive Cu-lowering therapeutic methods, mice injected with BRAF^{V600E}-transformed MEFs were provided with a Cu-deficient (CuD) diet supplemented with Cu and orally treated with vehicle alone for two weeks, or provided with a CuD diet and treated for two weeks with vehicle alone, oral TTM, or oral TTM followed by oral trientine (Fig. 4f). Tumours in mice fed with a CuD diet trended towards being smaller, although this did not reach significance ($P = 0.055$). Combining this diet with TTM blocked the development of visible tumours during the two weeks of treatment, although tumours quickly emerged after treatment ceased. However, mice treated with oral trientine after TTM showed a more durable response (Fig. 4g

Figure 4 | Pharmacological chelation of Cu decreases tumour growth of BRAF^{V600E}-driven and vemurafenib-resistant tumour cells. **a**, Normalized average soft agar growth (\pm s.e.m.) of *Ctr1*^{+/+} MEFs (plated in triplicate) expressing the indicated transgenes or the indicated cell lines treated with vehicle (black bar), 100 nM TTM (grey bar) or 400 nM TTM (white bar). Four asterisks, $P < 0.0001$. Results are representative of three experiments. **b, c, e**, Mean tumour volume (\pm s.e.m.) versus time for tumours of mice ($n = 3$) injected with BRAF^{V600E}-transformed *Ctr1*^{+/+} MEFs (**b**), DM440 cells (**c**) or ERK2^{GOF}-transformed *Ctr1*^{-/-} MEFs (**e**) and treated with either vehicle (black circles) or TTM (orange squares). Asterisk, $P < 0.05$; two asterisks, $P < 0.01$. **d**, Box-and-whiskers plot of percentage of P-ERK1/2 positive-stained cells per field of tumours from mice ($n = 3$) injected with BRAF^{V600E}-transformed *Ctr1*^{+/+} MEFs and treated with vehicle (black line) or TTM (orange line). Asterisk, $P < 0.05$. **f, g**, Dosing and diet regimen (**f**) and mean tumour volume (\pm s.e.m.) versus time (**g**) of mice ($n = 3$) injected with BRAF^{V600E}-transformed *Ctr1*^{+/+} MEFs and provided with a copper-deficient diet (CuD) with either deionized H₂O (diH₂O) and treated for two weeks with vehicle (group A, black circles) or diH₂O supplemented with CuSO₄ and treated for two weeks with vehicle (group B, blue triangles), TTM (group C, orange squares) or TTM then trientine (group D, green open diamonds). Three asterisks, $P < 0.001$. **h**, Immunoblot detection of the indicated mRNA and phosphorylated (P) or total (T) recombinant proteins with or without 50 μ M TTM. **i**, Immunoblot detection of the indicated recombinant proteins bound to resin charged with or without Cu. **j, k**, RT-PCR (**j**) and immunoblot detection of indicated phosphorylated (P) or total (T) endogenous, ectopic (ect), or both (end/ect) proteins (**k**) in A375 cells expressing haemagglutinin (HA)-tagged MEK1 or HA-MEK1^{C121S}. **l**, Mean tumour volume (\pm s.e.m.) versus time of mice injected with A375 cells expressing MEK1 (top) or MEK1^{C121S} mutant (bottom) and treated with vehicle (black circles, $n = 4$), TTM (orange squares, $n = 3$) or vemurafenib (blue triangles, $n = 4$). Results were compared using a one-tailed unpaired *t*-test (**b, c, d**) or a one-way ANOVA followed by a Tukey's multi-comparison post test (**a, g, l**). Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$. Gel images are representative of at least two technical replicates.

and Extended Data Fig. 10e). Thus, Cu-lowering strategies used to manage Wilson disease⁷ could be repurposed for the treatment of BRAF^{V600E}-driven cancers.

BRAF^{V600E} inhibitors such as vemurafenib have a limited duration of response as a result of acquired resistance that is often associated with the reactivation of MEK1/2 kinases²⁴. One reported mechanism of resistance is the acquisition of an activating C121S mutation²⁵ in MEK1. Although MEK1^{C121S} had elevated kinase activity *in vitro*, it still bound to the Cu-charged resin and its kinase activity was still inhibited by TTM (Fig. 4h, i). Moreover, although expressing MEK1^{C121S} in the BRAF mutation-positive melanoma cell line A375 increased P-ERK1/2 levels and imparted resistance to vemurafenib, in contrast with the same cells expressing MEK1, the MEK1^{C121S}-expressing A375 cells nevertheless remained sensitive to TTM *in vivo* (Fig. 4j–l and Extended Data Fig. 10f). Thus, Cu chelation decreases tumorigenicity driven by a mutation that confers resistance to a BRAF^{V600E} inhibitor.

Although the detailed mechanism underlying the requirement of Cu for robust MEK1/2 kinase activity remains to be explained, we have shown that decreasing the influx, bioavailability and binding of Cu to MEK1 decreases MEK1(2) kinase activity and oncogenic BRAF-driven tumorigenesis. Cu chelators, which are generally safe and economical drugs that have been given daily for decades to manage Cu levels in patients with Wilson disease⁷, also decreased oncogenic BRAF- and MEK1-driven tumorigenesis. Thus, inhibiting MEK1/2 kinase activity with Cu chelators, perhaps in combination with other MAPK pathway inhibitors, may merit clinical consideration for the treatment of not only BRAF mutation-positive cancers but also cancers developing resistance to BRAF^{V600E} and potentially even MEK1/2 inhibitors.

METHODS SUMMARY

Binding to Cu-charged resin, *in vitro* kinase assays, metal-catalysed oxidation, circular dichroism and differential scanning fluorimetry were performed as described previously^{6,26–28}. Standard methods were used to generate cell lines and assay RNA and protein levels and growth in soft agar. Measurements of tumour growth in immunocompromised mice treated with the indicated drugs and fed with the indicated diets as well as monitoring the tumour development and lifespan of BP mice²³ in a *Ctr1*^{+/+} or *Ctr1*^{flox/flox} background²² treated intranasally with AdCre²³ were performed under a protocol approved by the Institutional Animal Care and Use Committee.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 September 2012; accepted 24 February 2014.

Published online 9 April 2014.

1. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
2. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
3. Wan, P. T. *et al.* Mechanism of activation of the RAF–ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
4. Flaherty, K. T. *et al.* Improved survival with MEK inhibition in BRAF-mutated melanoma. *N. Engl. J. Med.* **367**, 107–114 (2012).
5. Flaherty, K. T. *et al.* Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N. Engl. J. Med.* **367**, 1694–1703 (2012).
6. Turski, M. L. *et al.* A novel role for copper in Ras/mitogen-activated protein kinase signaling. *Mol. Cell. Biol.* **32**, 1284–1295 (2012).
7. Ala, A., Walker, A. P., Ashkan, K., Dooley, J. S. & Schilsky, M. L. Wilson's disease. *Lancet* **369**, 397–408 (2007).
8. Solit, D. B. *et al.* BRAF mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–362 (2006).

9. Lee, J., Petris, M. J. & Thiele, D. J. Characterization of mouse embryonic cells deficient in the *ctr1* high affinity copper transporter. Identification of a *Ctr1*-independent copper transport system. *J. Biol. Chem.* **277**, 40253–40259 (2002).
10. Bertinato, J. & L'Abbe, M. R. Copper modulates the degradation of copper chaperone for Cu,Zn superoxide dismutase by the 26 S proteasome. *J. Biol. Chem.* **278**, 35071–35078 (2003).
11. Puig, S., Lee, J., Lau, M. & Thiele, D. J. Biochemical and genetic analyses of yeast and human high affinity copper transporters suggest a conserved mechanism for copper uptake. *J. Biol. Chem.* **277**, 26021–26030 (2002).
12. Bridgewater, J. D., Lim, J. & Vachet, R. W. Using metal-catalyzed oxidation reactions and mass spectrometry to identify amino acid residues within 10 Å of the metal in Cu-binding proteins. *J. Am. Soc. Mass Spectrom.* **17**, 1552–1559 (2006).
13. Fischmann, T. O. *et al.* Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. *Biochemistry* **48**, 2661–2674 (2009).
14. Huang, W., Kessler, D. S. & Erikson, R. L. Biochemical and biological analysis of MEK1 phosphorylation site mutants. *Mol. Biol. Cell* **6**, 237–245 (1995).
15. Nakamura, K., Uhlik, M. T., Johnson, N. L., Hahn, K. M. & Johnson, G. L. PB1 domain-dependent signaling complex is required for extracellular signal-regulated kinase 5 activation. *Mol. Cell. Biol.* **26**, 2065–2079 (2006).
16. English, J. M. *et al.* Contribution of the ERK5/MEK5 pathway to Ras/Raf signaling and growth control. *J. Biol. Chem.* **274**, 31588–31592 (1999).
17. Roberts, P. J. & Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310 (2007).
18. Levin-Salomon, V., Kogan, K., Ahn, N. G., Livnah, O. & Engelberg, D. Isolation of intrinsically active (MEK-independent) variants of the ERK family of mitogen-activated protein (MAP) kinases. *J. Biol. Chem.* **283**, 34500–34510 (2008).
19. Chu, Y., Solski, P. A., Khosravi-Far, R., Der, C. J. & Kelly, K. The mitogen-activated protein kinase phosphatases PAC1, MKP-1, and MKP-2 have unique substrate specificities and reduced activity *in vivo* toward the ERK2 sevenmaker mutation. *J. Biol. Chem.* **271**, 6497–6501 (1996).
20. Turski, M. L. & Thiele, D. J. New roles for copper metabolism in cell proliferation, signaling, and disease. *J. Biol. Chem.* **284**, 717–721 (2009).
21. Tsai, C. Y., Finley, J. C., Ali, S. S., Patel, H. H. & Howell, S. B. Copper influx transporter 1 is required for FGF, PDGF and EGF-induced MAPK signaling. *Biochem. Pharmacol.* **84**, 1007–1013 (2012).
22. Nose, Y., Kim, B. E. & Thiele, D. J. *Ctr1* drives intestinal copper absorption and is essential for growth, iron metabolism, and neonatal cardiac function. *Cell Metab.* **4**, 235–244 (2006).
23. Dankort, D. *et al.* A new mouse model to explore the initiation, progression, and therapy of BRAFV600E-induced lung tumors. *Genes Dev.* **21**, 379–384 (2007).
24. Solit, D. B. & Rosen, N. Resistance to BRAF inhibition in melanomas. *N. Engl. J. Med.* **364**, 772–774 (2011).
25. Wagle, N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J. Clin. Oncol.* **29**, 3085–3096 (2011).
26. Bridgewater, J. D. & Vachet, R. W. Metal-catalyzed oxidation reactions and mass spectrometry: the roles of ascorbate and different oxidizing agents in determining Cu-protein-binding sites. *Anal. Biochem.* **341**, 122–130 (2005).
27. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).
28. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**, 2212–2221 (2007).

Acknowledgements We thank M. McMahon, C. Cronin, E. Johnson, A. Stewart, D. S. Tyler and D. G. Kirsch for reagents, and C. Cronin, L. E. Crose, A. M. Jaeger, M. A. Luftig, E. Johnson, D. F. Kashatus, B. L. Lampson, J. P. Madigan, N. I. Nicely, Y. Nose, C. W. Pemble, N. L. K. Pershing, A. Stewart and J. D. Weyandt for technical support, discussions and/or review of the manuscript. This work was supported by National Institutes of Health grants CA178145 (D.C.B.), HL075443 (Proteomic Core K.X.), DK074192 (D.J.T.), CA094184 and CA172104 (C.M.C.), the Structural Genomics Consortium (Wellcome Trust 092809/Z/10/Z), FP7 grant 278568 'PRIMES' (S.K. and A.C.), the Stewart Trust (C.M.C.), the Edward Spiegel Fund of the Lymphoma Foundation (C.M.C.), and donations made in memory of Linda Woolfenden (C.M.C.).

Author Contributions Experiments were performed by D.C.B., M.S.C., M.L.T., G.A.H., X.Y. and K.X. All authors contributed to the study design. The manuscript was written by D.C.B. and C.M.C. with contributions by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.C. (chris.counter@duke.edu).

METHODS

Cell lines. *Ctr1*^{+/+} and *Ctr1*^{-/-} MEFs (immortalized with a plasmid encoding the SV40 T-Ag gene) were described previously⁹. A375 was purchased from ATCC. DM175, DM440, DM443, DM598, DM646, DM738 and DM792 cell lines were provided by D. S. Tyler²⁹. Mouse lung cancer cell lines were created and cultured as described previously³⁰. *Ctr1*^{+/+} and *Ctr1*^{-/-} immortalized MEFs and A375, DM175, DM646, DM440, DM598, DM738 and DM792 cell lines were stably infected with retroviruses derived from pBABE, pWZL or pSUPER-based vectors (see plasmids below) as described previously³¹. *Ctr1*^{+/+} MEFs were stably transfected with pCMV-based vectors (see plasmids below), using established protocols.

Plasmids. pBABEpuro-MYC-HIS-BRAF^{V600E}, pSUPER-retro-puro-tetO-RALB-scramble (encoding RALB scramble shRNA), pBABEpuro-CMYC^{T58A} and pCMV-HA-MEK1^{WT} were described previously^{6,32–34}. pBABEbleo-Flag-BRAF^{V600E} was created by PCR subcloning BRAF^{V600E} from the pBABEpuro-MYC-HIS-BRAF^{V600E} plasmid with primers designed to include an amino-terminal Flag tag. pWZLblasti-CTR1^{WT} and pWZLblasti-CTR1^{M154A} were created by PCR subcloning CTR1^{WT} or CTR1^{M154A} from the pCDNA3.1-CTR1^{WT} or pCDNA3.1-CTR1^{M154A} plasmids¹¹. pGEX6P1-GST-HA-MEK1^{WT} (GST, glutathione S-transferase) and pWZLblasti-HA-MEK1^{WT} were created by PCR subcloning MEK1^{WT} from the pENTR1-MEK1^{WT} plasmid³⁵ with primers designed to include an N-terminal haemagglutinin tag. pGEX6P1-GST-HA-MEK1^{CBM} and pWZLblasti-HA-MEK1^{CBM} (M187A/H188A/M230A/H239A; CBM), pGEX6P1-GST-HA-MEK1^{C121S}, pWZLblasti-HA-MEK1^{C121S}, pGEX6P1-GST-HA-MEK1^{M187A/H188A} and pGEX6P1-GST-HA-MEK1^{M230A/H239A} were created by introducing mutations corresponding to the indicated amino-acid changes by site-directed mutagenesis into MEK1^{WT} from the pENTR1-MEK1^{WT} plasmid³⁵ followed by PCR subcloning with primers designed to include an N-terminal HA tag. pGEX6P1-GST-HA-MEK5^{WT} and pWZLblasti-HA-MEK5^{WT} were created by PCR subcloning MEK5^{WT} from the pWZLneo-Myr-Flag-MEK5 plasmid³². pGEX6P1-GST-HA-MEK5^{CBM-E} and pWZLblasti-HA-MEK5^{CBM-E} (L280A/H281A/M323A/Q332A; CBM-E) were created by introducing mutations corresponding to the indicated amino-acid changes by site-directed mutagenesis into MEK5^{WT} from the pWZLneo-Myr-Flag-MEK5 plasmid³² followed by PCR subcloning with primers designed to include an N-terminal HA tag. pGEX6P1-GST-HA-ERK5^{K84R} was created by introducing a mutation corresponding to the indicated amino-acid change by site-directed mutagenesis into ERK5^{WT} from the pWZLneo-Myr-Flag-ERK5 plasmid³² followed by PCR subcloning with primers designed to include an N-terminal HA tag. pWZLblasti-HA-ERK5^{WT} was created by PCR subcloning ERK5^{WT} from the pWZLneo-Myr-Flag-ERK5 plasmid³² with primers designed to include an N-terminal HA tag. pBABEpuro-HA-MEK1-MEK5^{DD} and pWZLblasti-HA-MEK1-MEK5^{DD} (S311D/T315D; DD) were created by introducing mutations corresponding to the indicated amino-acid changes by site-directed mutagenesis into MEK5^{WT} from the pWZLneo-Myr-Flag-MEK5 plasmid³² followed by two-step PCR to fuse MEK1 (nucleotides 1–201) to MEK5 (nucleotides 498–1344) in frame with primers to include an N-terminal HA-tag. pGEX6P1-GST-HA-ERK2^{K54R}, pWZLblasti-HA-ERK2^{R67S}, pWZLblasti-HA-ERK2^{D321N}, pBABEpuro-HA-ERK2^{GOF} and pWZLblasti-HA-ERK2^{GOF} (R67S/D321N; GOF) were created by introducing mutations corresponding to the indicated amino acid changes by site-directed mutagenesis into ERK2^{WT} from the pDONR223-ERK2^{WT} plasmid³⁶ followed by PCR subcloning with primers designed to include an N-terminal HA tag. pSUPER-retro-puro-tetO-CTR1-shRNA was created to express the human *CTR1* shRNA target sequence 5'-AAAGCCAGCTTCTCTTTGG-3'. pSUPER-retro-puro-tetO-Mek1-shRNA was created to express the mouse *Mek1* shRNA target sequence 5'-GCCTCTCAGCTCATATGGAAT-3' (ref. 37). pCMV-HA-MEK1 encoding mutants M94A, H100A, H119A, M187A, H188A, C207A, M230A, H239A, M256A, M308A, C341A and H358A were created by site-directed mutagenesis of MEK1^{WT} from the pCMV-HA-MEK1^{WT} plasmid.

Reverse transcriptase PCR. RNA was purified from MEFs or tumour cell lines and reverse transcribed (RT) as described previously³⁰, and then PCR amplified with the primers 5'-ATCCTCATCAGCTCCCAATG-3' and 5'-CACATCACCA TGCCACTTTC-3' to detect human *BRAF*; 5'-CTGTTTCCGGTTTGGTGAT-3' and 5'-TGCCCAACAGTTTGTGTGT-3' to detect human *CTR1*; 5'-ATGAAC CATATGGGGATGAACCAT-3' and 5'-TCAATGGCAGTGCTCTGTGATG TC-3' to detect mouse *Ctr1*; 5'-GCACAGTCAAGGCCGAGAAT-3' and 5'-GC CTTCTCCATGGTG GTGAA-3' to detect mouse *Gapdh*; 5'-CCTTGAGGCCT TTCTTACCC-3' and 5'-CCCACGATGTACGGAGAGTT-3' to detect human *MEK1*; 5'-GTGAACACAGTGGGGAGAT-3' and 5'-CAGGAGGAGGAATG GGGTAT-3' to detect mouse *Mek1*; 5'-CCTTGCAAGAAGCTGGAG-3' and 5'-TCGGGACATGATGCTTTG-3' to detect human *MEK5*^{DD}; 5'-TG ATCACACAGGGTTCTCGA-3' and 5'-TGGAAGATGGGCCTGTTAG-3' to detect human *ERK2*; 5'-GAGAGACCTCACTGCTG-3' and 5'-GATGGTACA TGACAAGGTGC-3' to detect human *GAPDH*; and 5'-ACGAGACAAGCTCA CC-3' and 5'-TTTCCACACCTGGTTGC-3' to detect human *CMYC*. The fold change in the ratio of *CTR1* mRNA to total *GAPDH* mRNA was measured in ImageJ

software by boxing each band per representative image with the rectangular selection tool, and calculating the total area of the band in pixels. The total area of the *CTR1* mRNA band in pixels was normalized to the total area of the total *GAPDH* mRNA band in pixels. The fold change is shown in figures.

Immunoblot analysis. Equal quantities of lysates were isolated from the indicated cell lines after 24 h of serum starvation and then resolved by SDS-PAGE and immunoblotted as described previously⁶ with one of the following primary antibodies: mouse anti-MEK1, rabbit anti-ERK2, rabbit anti-MEK2, mouse anti-MEK1/2, rabbit anti-ERK1/2, rabbit anti-ERK5, rabbit anti-phospho(Ser 217/221)-MEK1/2, rabbit anti-phospho(Thr 202/Tyr 204)-ERK1/2, rabbit anti-phospho(Thr 218/Tyr 220)-ERK5 (Cell Signaling Technology), mouse anti-HA (Covance), rabbit anti-CCS (Santa Cruz Biotechnology), mouse anti-β-actin (Sigma), mouse anti-MEK5 (BD Transductions Laboratories), rat anti-mylenin basic protein (Millipore) or mouse anti-phospho(Thr 98)-MBP (Millipore), followed by detection with one of the following horseradish peroxidase-conjugated secondary antibodies: goat anti-rabbit IgG, goat anti-mouse IgG (Invitrogen), goat anti-mouse light-chain-specific IgG or mouse anti-rabbit light-chain-specific IgG (Jackson ImmunoResearch Laboratories) using enhanced chemiluminescence (ECL; GE Healthcare) or SuperSignal West Femto Chemiluminescent substrate (Pierce) detection reagents. The fold change in the ratio of phosphorylated protein to total protein was measured in ImageJ software by boxing each band per representative image with the rectangular selection tool, and calculating the total area of the band in pixels. The total area of the phosphorylated protein band in pixels was normalized to the total area of the total protein band in pixels. The average fold change is shown in figures.

Immunoprecipitation. Cell lysate (500 µg) was incubated overnight with 5 µl of mouse anti-HA antibody (Covance) at 4 °C. The immunoprecipitates were collected with GammaBind G Sepharose beads (GE Healthcare) for 2 h at 4 °C. SDS-PAGE analysis and immunoblotting were performed as described above.

Protein purification. Recombinant GST-ERK2, GST-MEK1, GST-ERK5 and GST-MEK5 proteins were expressed from pGEX-based vectors (see plasmids above) and purified as described previously⁶. Recombinant MEK1^{WT}, MEK1^{DD} (S218D/S222D; DD), MEK1^{CBM} and MEK1^{CBM/DD} proteins were a gift from A. Stewart, E. Johnson and C. Cronin. Recombinant GST-MEK2 and GST-MEK5 were purchased from Abnova. Recombinant MBP was purchased from Millipore.

In vitro kinase assays. MEK1, MEK2 and MEK5 *in vitro* kinase assays were performed as described previously⁶. In brief, 0.6 µg of recombinant GST-ERK2^{K54R}, GST-ERK5^{K84R} or MBP and 1.4 µg of recombinant MEK1, GST-MEK1, GST-MEK2 or GST-MEK5, or mutants thereof, were incubated for 30 min in 180 µl of kinase buffer in the presence or absence of 2.5 molar equivalents of CuSO₄, AgNO₃, FeNH₂SO₄, NiSO₄ or ZnSO₄ (Fig. 1h and Extended Data Fig. 1a), a seven-step titration of TTM (Sigma) from 0–50 µM (Fig. 1f), a fixed 50 µM concentration of TTM (Figs 1j, n and 4g and Extended Data Fig. 1c) and/or a fixed concentration of 2.5 µM CuSO₄ (Fig. 1n and Extended Data Fig. 1c) at 22 °C.

Copper binding. Recombinant GST-MEK1, GST-MEK2, GST-MEK5 proteins or mutants thereof (1 µg) were incubated for 2 h in 500 µl of RIPA containing 30 µl of Cu-pentadentate resin (Affiland) or 30 µl of free pentadentate resin (Affiland) at 4 °C. SDS-PAGE analysis and immunoblot were performed as described above. Cu pull-down experiments were performed as described previously⁶.

Metal-catalysed oxidation, mass spectrometry and data analysis. GST-MEK1 was loaded with Cu(II) by using a CuSO₄-histidine complex as described previously⁶. Metal-catalysed oxidation (MCO) reactions were performed as described previously²⁶. In brief, metal-catalysed oxidation (MCO) reactions were performed at 37 °C in 50 mM HEPES pH 8.0, 150 mM NaCl containing 40 µM GST-MEK1, 100 mM ascorbate and 1 mM H₂O₂. Untreated MEK1 was included as the control. The reactions were stopped by precipitating the proteins in methanol/chloroform followed by reduction for 30 min with 2.5 mM dithiothreitol (DTT) at 37 °C and alkylation with 20 mM iodoacetamide in the dark at room temperature for 30 min (refs 38, 39). In-solution digestion was performed overnight at 37 °C with trypsin (1:40 enzyme:protein ratio; modified, sequencing grade; Promega), which hydrolyses peptide bonds at the carboxyl end of lysine and arginine, or chymotrypsin (1:60 enzyme:protein ratio; Sigma), which hydrolyses peptide bonds at the carboxyl end of aromatic or large hydrophobic side chains of tyrosine, tryptophan, phenylalanine, methionine and leucine. The digested peptide samples were desalted with stage tips⁴⁰ and freeze-dried with a SpeedVac. Liquid chromatography–tandem mass spectrometry (LC–MS/MS) analyses were performed on a Thermo Scientific LTQ Orbitrap XL mass spectrometer (Thermo Scientific) with a Finnigan Nano-spray II electrospray ionization source. Digested peptides were injected onto a 75 µm × 150 mm BEH C₁₈ column (particle size 1.7 µm; Waters) and separated using a Waters nanoACQUITY Ultra Performance LC (UPLC) System (Waters). The LTQ Orbitrap XL was operated in data-dependent mode using the TOP10 strategy as described previously⁴¹. In brief, each scan cycle was initiated with a full MS scan of high mass accuracy (375–1,800 *m/z*; acquired in the Orbitrap XL at 6 × 10⁴ resolution setting and automatic gain control target of 10⁶), which was

followed by MS/MS scans (automatic gain control target 5,000; threshold 3,000) in the linear ion trap on the ten most abundant precursor ions. Selected ions were dynamically excluded for 30 s. Singly charged ions were excluded from MS/MS analysis. MS/MS spectra were searched by using the SEQUEST algorithm against the human MEK1 sequence. Search parameters allowed for two missed tryptic cleavages, a mass tolerance of ± 80 p.p.m., a static modification of 57.02146 Da (carboxyamidomethylation) on cysteine, and dynamic modifications of 15.99491 Da (oxidation) on methionine or histidine. The MS/MS spectra of matched peptides were validated manually. Modelling of the distance between the thiol groups of methionines or imidazole groups of histidines was performed with the three-dimensional structure of MEK1 in complex with magnesium and ATP- γ S determined at 2.1 Å (PDB 3EQD)¹³.

Circular dichroism spectroscopy. Circular dichroism (CD) data were collected on a JASCO J-815 CD spectrometer with a JASCO Peltier device and water bath to control the temperature. Experiments were performed in a 1-mm cuvette at a protein concentration of 5 μ M in 20 mM Tris-HCl pH 8.0 and 100 mM NaCl. Far-ultraviolet scans were collected from 200 nm to 250 nm. Thermal denaturation of MEK1 and MEK1^{CBM} proteins was monitored at 222 nm to estimate the protein melting temperature. The temperature ramp rate was 1 °C min⁻¹ and data points were collected every 1 °C. All data are reported in units of mean residue ellipticity, which was calculated from $[\theta]_{\text{MRE}} = (\theta_{\text{raw}} \times \text{MRW})/10cl$, where θ_{raw} is the ellipticity in degrees, MRW is molecular mass in daltons divided by (number of residues - 1), c is the protein concentration (g ml⁻¹), and l is the path length of the cuvette (cm), as described previously²⁷.

Differential scanning fluorimetry. Differential scanning fluorimetry data were collected on a CFX384 Touch™ Real-Time PCR Detection System (Bio-Rad) at a protein concentration of 1.3 μ M in 25 mM Tris-HCl pH 7.5, 20 mM MgCl₂, 2 mM DTT using SYPRO orange as described previously²⁸. The melting temperature was calculated by determining the maximum of the first-derivative curve of normalized data in Prism 5 (GraphPad) as described previously²⁸.

Mouse xenografts, drug treatments and diet alteration. MEFs (5×10^6) or melanoma cells (10^7) resuspended in PBS were injected subcutaneously into flanks of 8–12-week-old female SCID/beige mice (Charles River Laboratory) as described previously⁴². Unblinded drug treatments were as follows: vehicle (1% methylcellulose (Sigma), 1% dimethylsulphoxide (DMSO; Sigma)), 2.0 mg tetrathiomolybdate (TTM; Sigma) in vehicle, or 20 mg kg⁻¹ vemurafenib (Chemietek) in vehicle every other day by oral gavage. Mice were fed with a normal diet (PMI 5053 Picolab Mouse Diet 20; LabDiet) or, where indicated, a Cu-deficient diet (CuD diet, TD.80388; Harlan Teklad). Mice fed with a CuD diet were administered deionized H₂O (diH₂O) supplemented with 20 mg l⁻¹ CuSO₄, diH₂O alone, or diH₂O supplemented with 3 g l⁻¹ trientine dihydrochloride (Sigma). All studies were approved by the Duke University Institutional Animal Care and Use Committee. Statistical analysis of tumour volumes at endpoint was performed with a one-tailed, unpaired *t*-test, with a 95% confidence interval for two group data sets or one-way analysis of variance (ANOVA) with a 95%, 99% or 99.9% confidence interval and Tukey's multiple comparison post test for three or more data sets in Prism 5 (GraphPad). Statistical analysis of the percentage of mice with tumours of at least 1.0 cm³ versus time was analysed by using a survival curve log-rank (Mantel–Cox) test in Prism 5 (GraphPad).

Mouse lung cancer model. *Ctr1*^{fllox/fllox} mice²² were interbred with *Braf*^{CA/+}; *Trp53*^{fllox/fllox} (BP) mice²³, a gift from D. G. Kirsch, for three generations to generate *Braf*^{CA/+}; *Trp53*^{fllox/fllox}; *Ctr1*^{+/+} and *Braf*^{CA/+}; *Trp53*^{fllox/fllox}; *Ctr1*^{fllox/fllox} mice. PCR was performed to detect wild-type, conditional and recombined alleles of *Braf*, *Ctr1* and *Trp53*, as described previously^{22,23}. Cohorts of these animals of known genotype were administered 6×10^6 plaque-forming units of adenoviral Cre (University of Iowa) intranasally between 69 and 85 days of age. Mice were then monitored and killed euthanized three months later or at moribundity endpoints. All studies were approved by the Duke University Institutional Animal Care and Use Committee. Statistical analysis of percentage survival versus time was conducted with a survival curve log-rank (Mantel–Cox) test in Prism 5 (GraphPad).

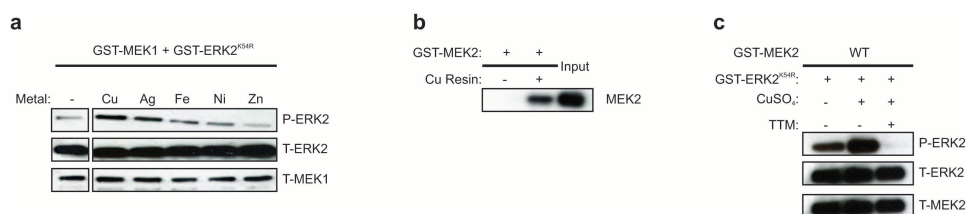
Analysis of lung tumours. To quantify the number of visible surface tumours per mouse, lungs were resected from five *Ctr1*^{+/+} and five *Ctr1*^{fllox/fllox} BP mice of known genotype three months after treatment with AdCre and the numbers of tumours visible on the surface of lungs were counted. Statistical analysis of the average number of tumours per mouse was performed using a one-tailed, unpaired *t*-test, with a 95% confidence interval for two group data sets. To quantify the percentage of abnormal lung tissue, lungs were resected from five *Ctr1*^{+/+} and five *Ctr1*^{fllox/fllox} BP mice three months after treatment with AdCre, fixed in 10% formalin and embedded in paraffin. Sections were deparaffinized, rehydrated and subjected to epitope retrieval before staining with haematoxylin (Surgipath) and eosin (Fisher Scientific). Photographs were taken at low power on an Olympus Vanox S microscope to encompass the entire lung; images were blinded, and the abnormal areas

of adenoma and adenocarcinoma from each micrograph were circumscribed using Adobe Photoshop and the abnormal area in pixels was expressed as a percentage of the total lung area from all micrographs. Statistical analysis of the average percentage of abnormal tissue per mouse was performed with a one-tailed, unpaired *t*-test, with a 95% confidence interval for two group data sets.

Immunohistochemistry. Three xenograft tumours from mice injected with BRAF^{V600E}-transformed *Ctr1*^{+/+} MEFs or lungs resected at moribundity endpoint from *Ctr1*^{+/+} and *Ctr1*^{fllox/fllox} BP mice treated with AdCre were fixed in 10% formalin and embedded in paraffin. Sections were deparaffinized, rehydrated and subjected to epitope retrieval and stained with an anti-P-ERK1/2 (Thr 202/Tyr 204) antibody (Cell Signaling), followed by peroxidase-based detection with Vectastain Elite ABC Kits (Vector Labs) and counterstaining with haematoxylin. Photographs were taken of high-power fields of highest positivity (P-ERK IHC) on an Olympus Vanox S microscope. Images were blinded and tumours were circumscribed in P-ERK1/2-stained tissue images with the freehand selection tool in Image J, and the total area of the tumour in pixels was recorded. Tumour images were copied and pasted to new, blank images and colour thresholding was applied to determine positive-staining areas, using the same parameters for each tumour image. Areas staining positive by these parameters were selected and the positive-staining area in pixels was recorded. The percentage positive-staining area was calculated by dividing the positive-staining area of the tumour in pixels by the total area of the tumour in pixels. Statistical significance was determined with unpaired, one-tailed *t*-tests between treated and untreated, or between *Ctr1*^{+/+} and *Ctr1*^{fllox/fllox} cohorts, using Prism 5 (GraphPad).

Cell growth and soft agar assays. For growth-curve experiments, cells were plated at a density of 5,000 cells per well in six replicate wells in four 24-well plates. At days 0, 1, 2 and 3, one plate was fixed for 5 min in formalin, washed with PBS and stained for 30 min with 0.1% crystal violet (Sigma) and left to dry for at least 24 h. Stain was extracted in 200 μ l of 10% acetic acid in each well, transferred to a 96-well plate, and the absorbance at 600 nm was measured with a GloMax Multi Detection System plate reader (Promega) in an unblinded fashion. Relative growth was determined by normalization to the signal at day 0 and plotted in Prism 5 (GraphPad). Anchorage-independent growth was assayed in six-well plates with 1 ml of 0.6% bactoagar media solution (final concentration $1 \times$ DMEM, 10% FBS, $1 \times$ penicillin/streptomycin) as a bottom support layer; 5×10^4 cells per well were resuspended in DMEM (10% FBS, $1 \times$ penicillin/streptomycin) and mixed 1:1 with 0.6% bactoagar media solution with an appropriate concentration of DMSO or TTM to give a final bactoagar concentration of 0.3%, 100 nM TTM, 400 nM TTM, or an equivalent amount of DMSO and plated in triplicate. Each well was fed on days 3, 7, 14 and 21 with 300 μ l of DMEM with the appropriate concentration of DMSO or TTM. Colonies were counted unblinded between days 21 and 28. Statistical analysis of soft agar growth was performed with a two-way ANOVA with a 99.9% confidence interval and Tukey's multiple comparison post test in Prism 5 (GraphPad).

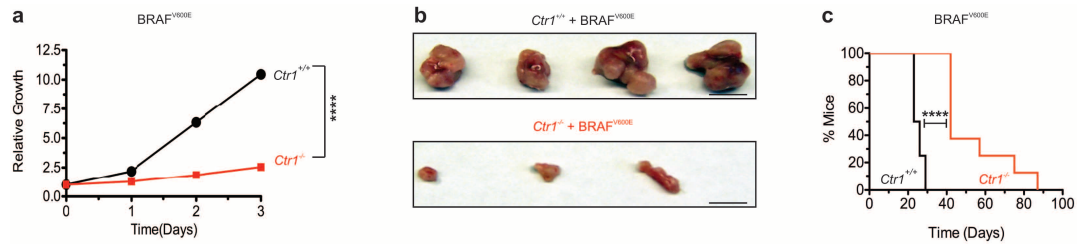
29. Augustine, C. K. *et al.* Sorafenib, a multikinase inhibitor, enhances the response of melanoma to regional chemotherapy. *Mol. Cancer Ther.* **9**, 2090–2101 (2010).
30. Lampson, B. L. *et al.* Targeting eNOS in pancreatic cancer. *Cancer Res.* **72**, 4472–4482 (2012).
31. O'Hayer, K. M. & Counter, C. M. A genetically defined normal human somatic cell system to study Ras oncogenesis *in vivo* and *in vitro*. *Methods Enzymol.* **407**, 637–647 (2006).
32. Boehm, J. S. *et al.* Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell* **129**, 1065–1079 (2007).
33. O'Hayer, K. M., Brady, D. C. & Counter, C. M. ELR+ CXCL chemokines and oncogenic Ras-mediated tumorigenesis. *Carcinogenesis* **30**, 1841–1847 (2009).
34. Zipfel, P. A. *et al.* Ral activation promotes melanomagenesis. *Oncogene* **29**, 4859–4864 (2010).
35. Scholl, F. A., Dumesic, P. A. & Khavari, P. A. Mek1 alters epidermal growth and differentiation. *Cancer Res.* **64**, 6035–6040 (2004).
36. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
37. Bric, A. *et al.* Functional identification of tumor-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16**, 324–335 (2009).
38. Kahsai, A. W. *et al.* Multiple ligand-specific conformations of the β_2 -adrenergic receptor. *Nature Chem. Biol.* **7**, 692–700 (2011).
39. Nobles, K. N. *et al.* Distinct phosphorylation sites on the β_2 -adrenergic receptor establish a barcode that encodes differential functions of beta-arrestin. *Sci. Signal.* **4**, ra51 (2011).
40. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
41. Haas, W. *et al.* Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **5**, 1326–1337 (2006).
42. Hamad, N. M. *et al.* Distinct requirements for Ras oncogenesis in human versus mouse cells. *Genes Dev.* **16**, 2045–2057 (2002).



Extended Data Figure 1 | CuSO₄ stimulates MEK1/2 kinase activity *in vitro*.

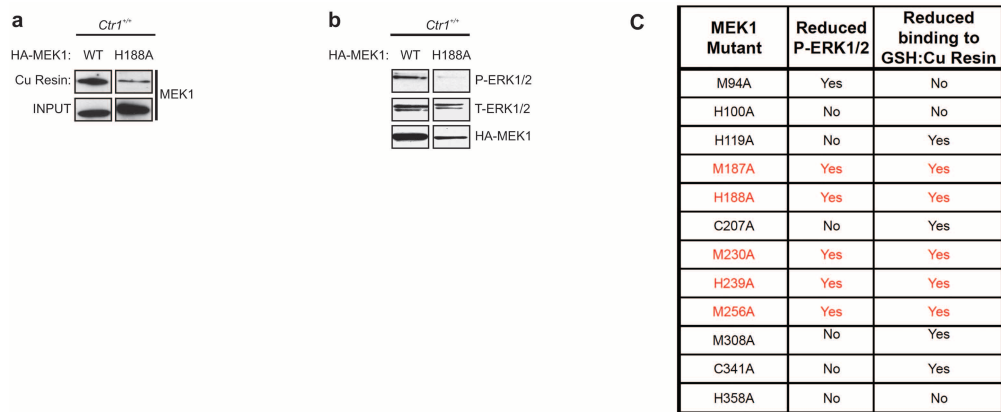
a, c, Detection of the amount of *in vitro* phosphorylated (P) recombinant GST-tagged kinase-inactive ERK2^{K54R} protein by recombinant GST-tagged MEK1 in the presence, when indicated, of 2.5 molar equivalents of CuSO₄ (Cu), AgNO₃ (Ag), FeNH₂SO₄ (Fe), NiSO₄ (Ni) or ZnSO₄ (Zn) (**a**) or recombinant GST-tagged MEK2 in the presence, when indicated, of 2.5 μM CuSO₄ and/or

50 μM TTM (**c**). Total (T) levels of ERK2, MEK1 and MEK2 serve as loading controls. **b,** Immunoblot detection of the amount of recombinant GST-tagged MEK2 protein bound to a resin charged with (Cu) or without (–) Cu. Input serves as a loading control. Gel images are representative of two technical replicates.



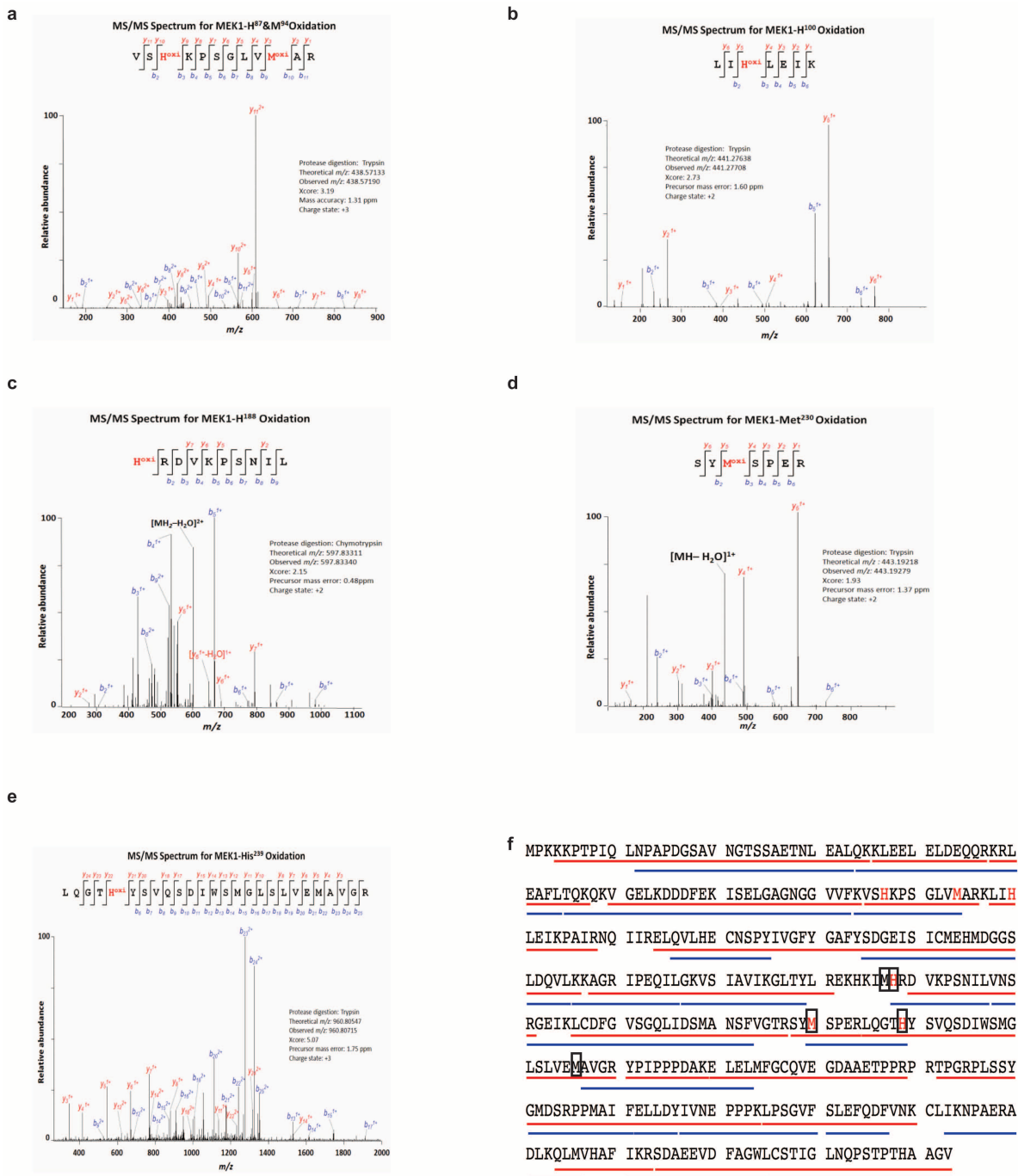
Extended Data Figure 2 | Genetic ablation of *Ctrl* decreases BRAF^{V600E}-mediated cell growth and tumorigenesis. **a**, Cell growth, as measured by staining with crystal violet, of BRAF^{V600E}-transformed *Ctrl*^{+/+} (black circles) or *Ctrl*^{-/-} (red squares) MEFs (plated in sextuplicate) over a period of three days. Representative of three independent experiments using the same cells. **b, c**, Representative resected tumours (scale bar, 1 cm) at 20 days after injection

(**b**) and Kaplan–Meier analysis of percentage of mice with tumour volume at least 1.0 cm³ versus time (**c**) of mice ($n = 8$) injected with BRAF^{V600E}-transformed *Ctrl*^{+/+} (black line) or *Ctrl*^{-/-} (red line) MEFs. Results were compared using a one-tailed unpaired *t*-test (**a**) or a Mantel–Cox test (**c**). Four asterisks, $P < 0.0001$.



Extended Data Figure 3 | Identification of Cu-binding mutants of MEK1 that decrease ERK1/2 phosphorylation. **a**, Immunoblot detection of the amount of HA-tagged wild-type MEK1 and an example of one MEK1 mutant tested (H188A) that bound to a Cu-charged resin. Input serves as a loading control. **b**, Immunoblot detection of the amount of phosphorylated (P) and/or total (T) ERK1/2 or HA-MEK1 protein in *Ctrl^{+/+}* MEFs stably expressing

HA-tagged wild-type MEK1 or an example of one MEK1 mutant tested (H188A). **c**, Summary of whether the indicated MEK1 point mutants did (YES) or did not (NO) show a decrease in binding to the Cu-charged resin or show a decrease in the levels of phosphorylated (P) ERK1/2 when stably expressed in *Ctrl^{+/+}* MEFs. Gel images are representative of two technical replicates.



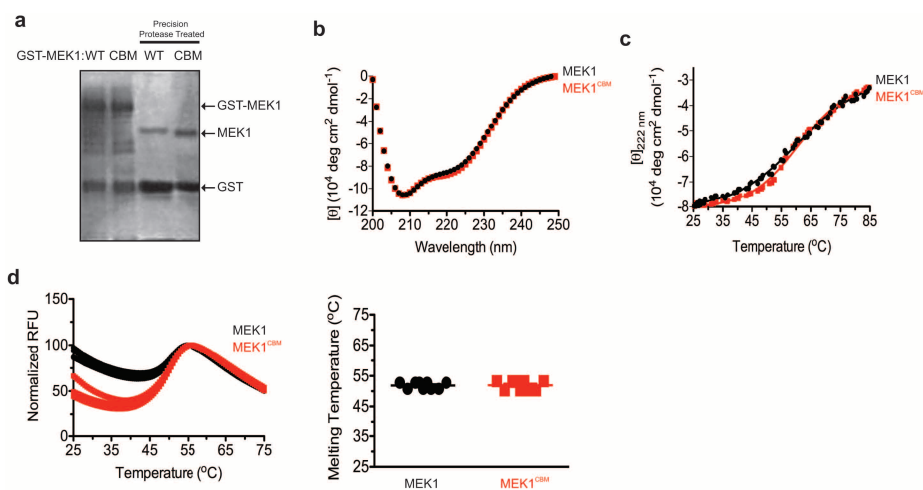
Extended Data Figure 4 | Amino acids in MEK1 identified to be oxidized by the MCO reaction followed by MS/MS. Representative annotated MS/MS fragmentation spectra for five indicated MEK1-derived peptides containing oxidized residues highlighted in red: **a**, MEK1^{H87} and MEK1^{M94}; **b**, MEK1^{H100}; **c**, MEK1^{H188}; **d**, MEK1^{K230}; **e**, MEK1^{H239}. The peak heights are the relative abundances of the corresponding fragmentation ions, with the annotation of the identified matched amino-terminus-containing ions (*b* ions) in blue and the carboxy-terminus-containing ions (*y* ions) in red. For clarity, only the

major identified peaks are labelled. **f**, Amino-acid sequence of human MEK1 with the peptides identified by MS/MS underlined (red, trypsin digest; blue, chymotrypsin digest). Amino acids oxidized only in the presence of H_2O_2 in one to three independent MCO reactions are denoted in red. Boxes enclose amino acids that when mutated to alanine decreased both the binding of MEK1 to a Cu-charged resin and the phosphorylation of cellular ERK2 (from Extended Data Fig. 3c).

MEK1	-----MPKKKPTP--IQLNP-----	13
MEK2	-----MLARRKPVLPALTINPT-----	17
MEK5	MLWLALGPPFAMENQVLVIRIKIPNSGAVDWTVHSGPQLLFRDVLVDVIGQVLPEATTTAF	60
MEK1-MEK5	-----MPKKKPTPIQLNPAP-----	15
MEK1	---APDGSVNGTSSAETN-----LEALQ-----KKLEEL	40
MEK2	---IAEGPSPTSEGASEAN-----LVDLQ-----KKLEEL	44
MEK5	EYEDGDGRITVRSDEEMKAMLSYYYSTVMEQQVNGQLIEPLQIFPRACKPPGERNIHGL	120
MEK1-MEK5	-----DGSVNGTSSAETN-----LEALQ-----KKLEEL	40
MEK1	ELDEQQR-----KRLEAFLTQKQKVGELKDDDFEKISELGAGNGGVV	82
MEK2	ELDEQQK-----KRLEAFLTQKAKVGELKDDDFERISELGAGNGGVV	86
MEK5	KVNTRAGPSQHSSPAVSDSLPSNSLKKSSAELKKILANGQMNEQDIRYRDTLGHGNGGTV	180
MEK1-MEK5	ELDEQQR-----KRLEAFLTQKQKVGELKDDDIRYRDTLGHGNGGTV	82
MEK1	FKVSHKPSGLVMARKLIHLEIKPAIRNQIIRELQVLHECNSPYIVGFYGAFFSDGEISIC	142
MEK2	TKVQHRPSGLIMARKLIHLEIKPAIRNQIIRELQVLHECNSPYIVGFYGAFFSDGEISIC	146
MEK5	YKAYHVPSGKILAVKVILLDTITLELQKQIMSELEILYKCDSSYIIGFYGAFFVENRISIC	240
MEK1-MEK5	YKAYHVPSGKILAVKVILLDTITLELQKQIMSELEILYKCDSSYIIGFYGAFFVENRISIC	142
MEK1	MEHMDGGSGLDQVLKAGRIPEQILGKVSIAVIKGLTYLREKHKIMHRDVKPSNMLVNSRG	202
MEK2	MEHMDGGSGLDQVLKEAKRIPEEILGKVSIAVLRGLAYLREKHQIMHRDVKPSNMLVNSRG	206
MEK5	TEFMDGGSGLDVYR---KMPEHVLGRIAVAVVGLTYLWS-LKILHRDVKPSNMLVNTRG	295
MEK1-MEK5	TEFMDGGSGLDVYR---KMPEHVLGRIAVAVVGLTYLWS-LKILHRDVKPSNMLVNTRG	197
MEK1	EIKLCDFGVSGQLIDSMANSFVGTRSYMSPERLQGTTHYSVQSDIWSMGLSLVEMAVGRYP	262
MEK2	EIKLCDFGVSGQLIDSMANSFVGTRSYMAPERLQGTTHYSVQSDIWSMGLSLVELAVGRYP	266
MEK5	QVKLCDFGVSTQLVNSIAKTYVGTNAYMAPERISGEQYGIHSDVWSLGISFMELALGRFP	355
MEK1-MEK5	QVKLCDFGVSTQLVNSIAKTYVGTNAYMAPERISGEQYGIHSDVWSLGISFMELALGRFP	257
MEK1	IPPPDAKELELMFG-CQVEGDAAETP---PRPRTPGRPLSSYGMDSRPPMAIFELLDYIV	318
MEK2	IPPPDAKELEAIFGRPVVDGEEGEPHSISPRPRPPGRPVSGHGMDSRPAMAIFFELLDYIV	326
MEK5	YP-----QIQKNQG-----SLMPLQLLQCIV	376
MEK1-MEK5	YP-----QIQKNQG-----SLMPLQLLQCIV	278
MEK1	NEPPPKLPNGVFTPDFQEFVNKCLIKNPAERADLKQMLVHAFIKRS-DAEEVDFAGWLCS	377
MEK2	NEPPPKLPNGVFTPDFQEFVNKCLIKNPAERADLKMLTNHTFIKRS-EVEEVDFAGWLCK	385
MEK5	DEDSPVLPVGEFSEPFVHFITQCMRKQPKERPAPPEELMGHPFIVQFNDGNAAVSMWVCR	436
MEK1-MEK5	DEDSPVLPVGEFSEPFVHFITQCMRKQPKERPAPPEELMGHPFIVQFNDGNAAVSMWVCR	338
MEK1	TIG--LNQPSTPTHAAGV	393
MEK2	TLR--LNQPGTPTRTAV-	400
MEK5	ALEERRSQQGPP-----	448
MEK1-MEK5	ALEERRSQQGPP-----	350

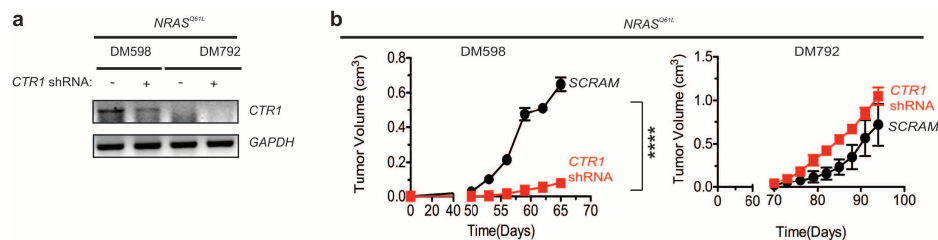
Extended Data Figure 5 | Alignment of the amino-acid sequences of MEK1, MEK2, MEK5 and MEK1-MEK5. The amino-acid sequences of human MEK1, MEK2, MEK5 and the MEK1-MEK5 chimaeric protein (without the DD mutation) aligned using Clustal W. Black letters, amino acids; coloured

letters, the four amino acids mutated in MEK1^{CBM} to decrease Cu binding (blue, conserved between MEK1, MEK2 and MEK5; red, conserved only between MEK1 and MEK2). Dashes indicate gaps in the alignment.



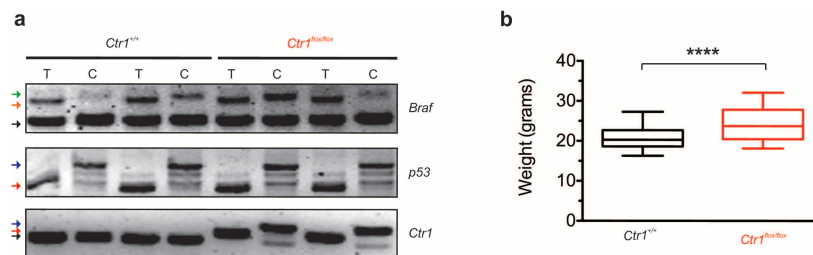
Extended Data Figure 6 | Protein purification and biochemical analysis of wild-type and CBM versions of MEK1. **a**, Coomassie blue detection of the amount of wild-type or CBM mutant purified recombinant GST-tagged MEK1 protein in the absence or presence of precision protease for cleavage of GST. **b–d**, Circular dichroism spectra at increasing wavelengths (**b**), thermal

denaturation monitored at 222 nm at increasing temperature (**c**), and differential scanning fluorimetry at increasing temperature (**d**, left) and the average estimated melting temperature (**d**, right) of purified recombinant MEK1^{WT} (black circles and line) and MEK1^{CBM} (red squares and line). Data are representative of two technical replicates.



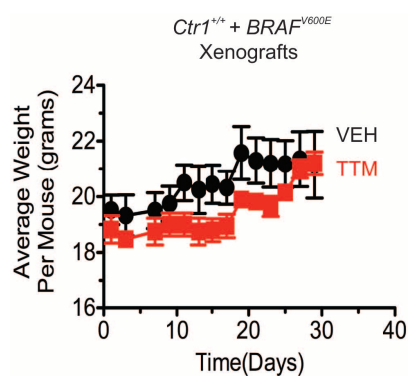
Extended Data Figure 7 | Tumorigenic growth of *NRAS* mutation-positive human melanoma cancer cell lines on knockdown of *CTR1*. **a, b**, RT-PCR detection of the amount of endogenous *CTR1* and *GAPDH* mRNA (**a**) and mean tumour volume (\pm s.e.m.) versus time (**b**) of mice ($n = 3$) injected with

the *NRAS* mutation-positive (*NRAS*^{Q61L}) human melanoma cell lines DM598 and DM792 stably infected with a retrovirus expressing either a *scramble* (*SCRAM*) shRNA (black circles) or *CTR1* shRNA (red squares). Results were compared using a one-tailed unpaired *t*-test (**b**). Four asterisks, $P < 0.0001$.

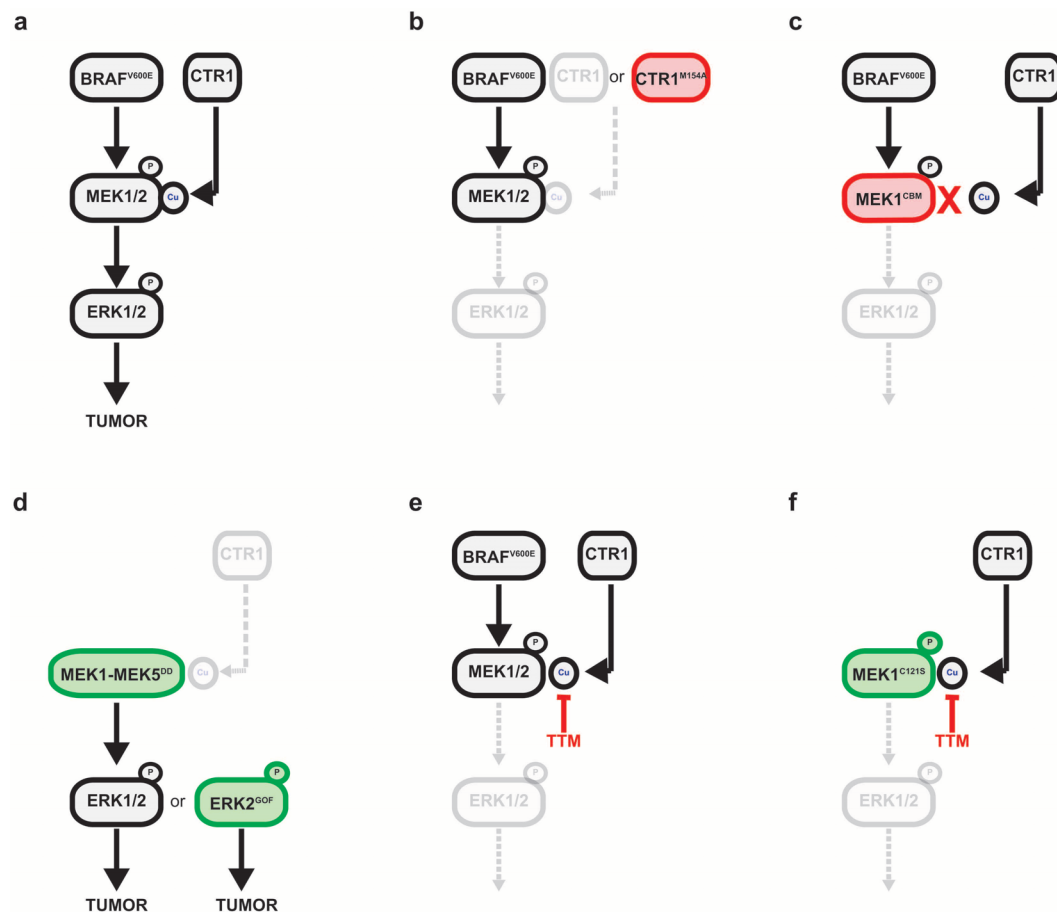


Extended Data Figure 8 | Detection of Cre-mediated recombination and weight measurements of AdCre-treated *Ctr1*^{+/+} versus *Ctr1*^{flox/flox} BP mice. **a**, PCR detection of *Braf*^{CA/+}, *Trp53*^{flox/flox} and *Ctr1*^{flox/flox} recombined alleles from matched tail samples (T) and lung tumour cell lines (C) generated from indicated genotypes. Alleles are indicated by arrowheads as follows: black,

WT; red, *flox*; blue, null; orange, *Braf*^{CA}; green, *Braf*^{V600E}. **b**, Box-and-whiskers plot of weight of *Ctr1*^{+/+} versus *Ctr1*^{flox/flox} BP mice ($n = 30$) one month after intranasal treatment with AdCre. Results were compared using a one-tailed unpaired *t*-test (**b**). Four asterisks, $P < 0.0001$.



Extended Data Figure 9 | TTM does not decrease the weight of mice with tumours. Mean weight (\pm s.e.m.) over time of mice ($n = 4$) injected with BRAF^{V600E}-transformed MEFs and treated with vehicle (black circles) or TTM (red squares).



Extended Data Figure 10 | Graphical representation of Cu regulation of BRAF^{V600E}-mediated signalling and tumorigenesis. Inactivation of the signalling pathway is denoted in grey and dashed lines, gain-of-function

mutations are denoted in green, and loss-of-function mutations are denoted in red.

Protective mucosal immunity mediated by epithelial CD1d and IL-10

Torsten Olszak^{1*}, Joana F. Neves^{1*}, C. Marie Dowds^{2*}, Kristi Baker¹, Jonathan Glickman³, Nicholas O. Davidson⁴, Chyuan-Sheng Lin⁵, Christian Jobin⁶, Stephan Brand⁷, Karl Sotlar⁸, Koichiro Wada⁹, Kazufumi Katayama⁹, Atsushi Nakajima¹⁰, Hiroyuki Mizuguchi¹¹, Kunito Kawasaki¹², Kazuhiro Nagata¹², Werner Müller¹³, Scott B. Snapper^{1,14}, Stefan Schreiber², Arthur Kaser¹⁵, Sebastian Zeissig^{1,2*} & Richard S. Blumberg^{1*}

The mechanisms by which mucosal homeostasis is maintained are of central importance to inflammatory bowel disease. Critical to these processes is the intestinal epithelial cell (IEC), which regulates immune responses at the interface between the commensal microbiota and the host^{1,2}. CD1d presents self and microbial lipid antigens to natural killer T (NKT) cells, which are involved in the pathogenesis of colitis in animal models and human inflammatory bowel disease^{3–8}. As CD1d crosslinking on model IECs results in the production of the important regulatory cytokine interleukin (IL)-10 (ref. 9), decreased epithelial CD1d expression—as observed in inflammatory bowel disease^{10,11}—may contribute substantially to intestinal inflammation. Here we show in mice that whereas bone-marrow-derived CD1d signals contribute to NKT-cell-mediated intestinal inflammation, engagement of epithelial CD1d elicits protective effects through the activation of STAT3 and STAT3-dependent transcription of IL-10, heat shock protein 110 (HSP110; also known as HSP105), and CD1d itself. All of these epithelial elements are critically involved in controlling CD1d-mediated intestinal inflammation. This is demonstrated by severe NKT-cell-mediated colitis upon IEC-specific deletion of IL-10, CD1d, and its critical regulator microsomal triglyceride transfer protein (MTP)^{12,13}, as well as deletion of HSP110 in the radioresistant compartment. Our studies thus uncover a novel pathway of IEC-dependent regulation of mucosal homeostasis and highlight a critical role of IL-10 in the intestinal epithelium, with broad implications for diseases such as inflammatory bowel disease.

To examine the function of CD1d in the intestinal epithelium, we generated mice with tamoxifen-inducible IEC-specific deletion of MTP, a lipid transfer protein encoded by *Mtpp* and required for CD1d function^{12,13} (Villin-CreER^{T2} (ref. 14) \times *Mtpp*^{fl/fl} (ref. 15); hereafter referred to as *Mtpp*^{ΔIEC}). *Mtpp*^{ΔIEC} mice exhibited no evidence of epithelial lipid accumulation (Extended Data Fig. 2a) and endoplasmic reticulum (ER) stress (Extended Data Fig. 2b), but showed decreased IEC-mediated, CD1d- but not major histocompatibility complex (MHC)-class-I-dependent antigen presentation despite unimpaired CD1d expression (Extended Data Fig. 3a, b, c left). *Mtpp*^{ΔIEC} mice exhibited increased sensitivity to oxazolone-induced colitis, a CD1d- and NKT-cell-dependent colitis model⁴, as shown by increased mortality (Fig. 1a, left), weight loss (Fig. 1a, right) and histological injury (Fig. 1b and Extended Data Fig. 3d), in comparison with wild-type littermates (Villin-CreER^{T2}-negative *Mtpp*^{fl/fl} mice). Increased severity of colitis was not caused by primary epithelial barrier defects (Extended Data Fig. 4a) but was due to loss of function

of intestinal epithelial CD1d. Thus, bone marrow chimaeras with selective deficiency of CD1d in the radioresistant compartment exhibited severe mortality and morbidity in the oxazolone model (Fig. 1c, d), increased production of pathogenic IL-13 and IL-1 β in colon explants (Fig. 1e) and purified CD11b⁺ lamina propria cells (Extended Data Fig. 4b), and reduced secretion of protective IL-10 (Fig. 1e). In accordance with a protective role of intestinal epithelial CD1d, its expression was upregulated in response to oxazolone colitis in wild-type, but not *Mtpp*^{ΔIEC} mice (Extended Data Fig. 3c, right). Intriguingly, pathogenic signals were similarly CD1d-dependent but emanated from non-epithelial cells in a manner dependent on adaptive immune cells. Thus, *Mtpp*^{ΔIEC} mice on a *Rag1*^{-/-} or *Cd1d1*^{-/-} *Cd1d2*^{-/-} (CD1d-knockout) background were protected from oxazolone colitis (Extended Data Fig. 4c, d). Consistent with a role of non-epithelial cells in the delivery of pathogenic CD1d signals, antibody-mediated neutralization of (non-epithelial) CD1d in *Mtpp*^{ΔIEC} mice reduced the severity of oxazolone-induced colitis (Fig. 1f, g). Thus, intestinal epithelial CD1d- and MTP-dependent lipid antigen presentation protects from oxazolone colitis, whereas CD1d- and NKT-cell-mediated inflammation emanates from non-epithelial cells, presumably bone-marrow-derived lamina propria cells.

To investigate directly the role of epithelial CD1d, we developed mice with inducible, IEC-specific deletion of CD1d (Villin-CreER^{T2} \times *Cd1d1*^{fl/fl}; hereafter referred to as *Cd1d1*^{ΔIEC}) (Extended Data Fig. 5a). Tamoxifen-treated *Cd1d1*^{ΔIEC} mice exhibited normal numbers of invariant (i)NKT cells and lacked intestinal epithelial expression of CD1d and CD1d- but not MHC-class-I-mediated antigen presentation by IECs (Extended Data Fig. 5b–e). Oxazolone-challenged *Cd1d1*^{ΔIEC} mice, like *Mtpp*^{ΔIEC} mice, exhibited increased mortality (Fig. 1h, left), weight loss (Fig. 1h, right), histopathological injury (Fig. 1i), colon shortening (Extended Data Fig. 5f) and elevated secretion of IL-13 and IL-1 β , as well as impaired secretion of IL-10 from colon explant cultures (Fig. 1j). Moreover, colonic iNKT cells from *Cd1d1*^{ΔIEC} mice showed increased IL-13 and IFN- γ expression in oxazolone colitis (Fig. 1k). Thus, intestinal epithelial CD1d protects from intestinal inflammation.

To gain insight into the mechanisms underlying protection by intestinal epithelial CD1d, we performed transcriptional profiling of intestinal mucosal scrapings from wild-type and *Mtpp*^{ΔIEC} mice in the oxazolone model (Extended Data Table 1 and data not shown). HSP110, encoded by *Hsp110*, exhibited decreased expression in *Mtpp*^{ΔIEC} mice (Fig. 2a), which was of interest given its known role in induction of epithelial CD1d expression¹⁶. In line with a potential MTP- and CD1d-dependent role of

¹Division of Gastroenterology, Hepatology, and Endoscopy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Internal Medicine I, University Medical Center Schleswig-Holstein, 24105 Kiel, Germany. ³GI Pathology, Miraca Life Sciences, Newton, Massachusetts 02464, USA. ⁴Division of Gastroenterology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁵Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York 10032, USA. ⁶Department of Medicine, Department of Infectious Diseases & Pathology, University of Florida, Gainesville, Florida 32611, USA. ⁷Department of Medicine II-Grosshadern, Ludwig Maximilians University, Munich 81377, Germany. ⁸Institute of Pathology, Ludwig Maximilians University, Munich 80337, Germany. ⁹Department of Pharmacology, Graduate School of Dentistry, Osaka University, Osaka 565-0871, Japan. ¹⁰Gastroenterology Division, Yokohama City University School of Medicine, Yokohama, Kanagawa 236-0027, Japan. ¹¹Laboratory of Biochemistry and Molecular Biology, Graduate School of Pharmaceutical Sciences, Osaka University, Osaka 565-0871, Japan. ¹²Department of Molecular Biosciences, Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan. ¹³Faculty of Life Sciences, University of Manchester, Manchester M13 9PL, UK. ¹⁴Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Department of Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. ¹⁵Division of Gastroenterology, Addenbrooke Hospital, University of Cambridge, Cambridge CB2 0QQ, UK.

*These authors contributed equally to this work.

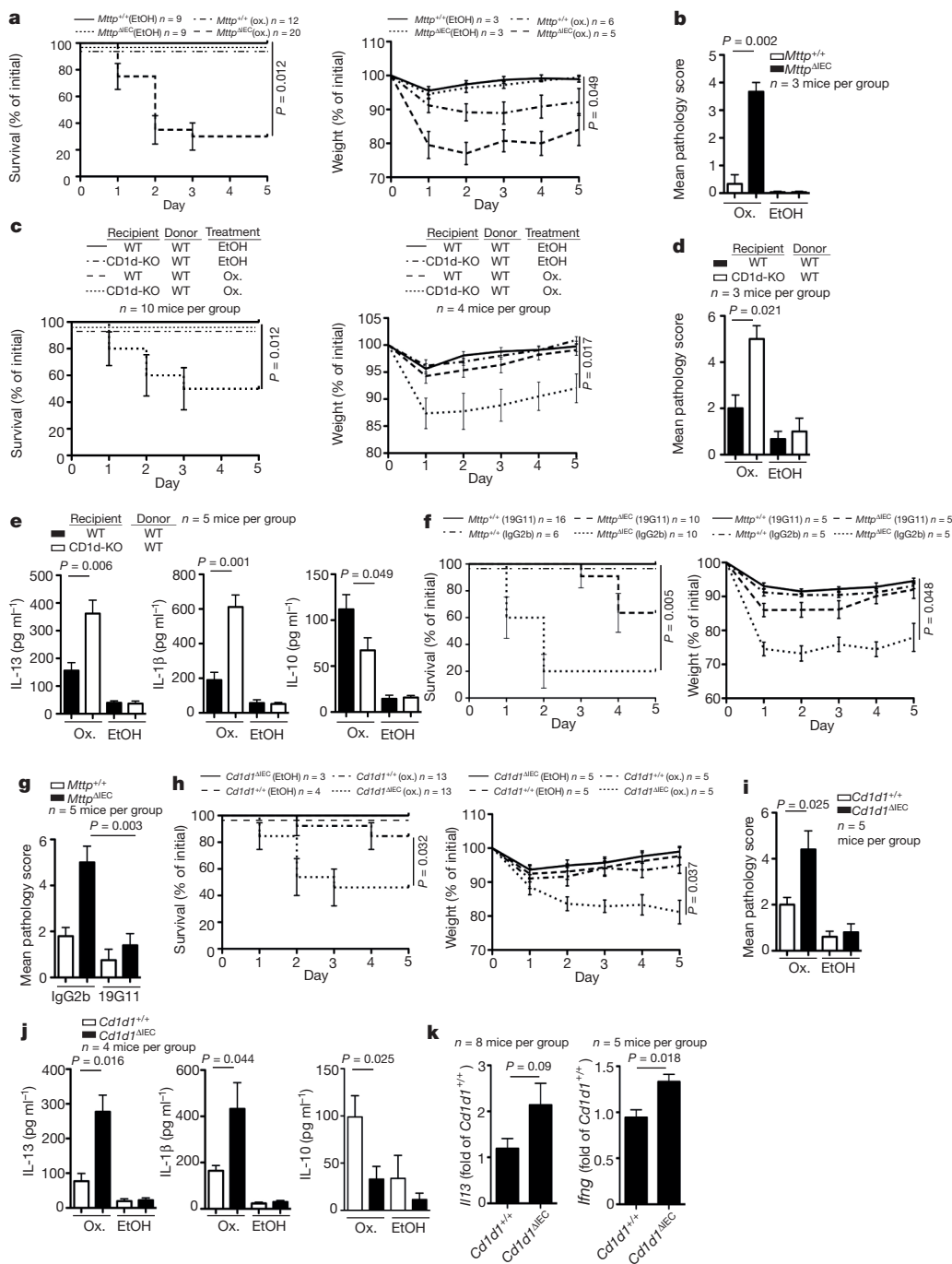


Figure 1 | Intestinal epithelial MTP and CD1d protect from oxazolone colitis. **a–k**, Survival (**a**, left, **c**, left, **f**, left, **h**, left), body weight (**a**, right, **c**, right, **f**, right, **h**, right), histopathology (**b**, **d**, **g**, **i**), cytokine secretion (enzyme-linked immunosorbent assay (ELISA)) in colon explant cultures (**e**, **j**), and cytokine expression (quantitative polymerase chain reaction (qPCR)) of sorted colonic iNKT cells combined from intraepithelial and lamina propria compartments (**k**) of the indicated mice upon rectal challenge with oxazolone (ox., **a–k**) or vehicle (ethanol (EtOH, **a–e**, **h–j**)). **c–e**, Wild-type (WT) and CD1d-knockout (CD1d-KO) mice were reconstituted with wild-type bone marrow. **f**, **g**, 19G11 (anti-CD1d) or isotype control antibody (IgG2b) were administered. Results representative of three independent experiments are shown. Mean \pm standard error of the mean (s.e.m.) of the indicated number of mice is shown. **j**, Right, Mann–Whitney *U*-test was applied. In all other panels, Student's *t*-test or log-rank test (survival) were applied.

HSP110 in intestinal inflammation, HSP110 messenger RNA (Fig. 2b, left) and protein expression (Extended Data Fig. 6a) were increased in IECs from wild-type but not *Mtpp*^{ΔIEC} mice upon oxazolone challenge. Similar observations were made for *Cd1d1*^{ΔIEC} mice (Fig. 2b, right). In addition, intestinal epithelial but not lamina propria HSP110 expression was reduced in active and inactive human ulcerative colitis compared with healthy controls (Fig. 2c and Extended Data Fig. 6b), which is in accordance with impaired epithelial CD1d expression in inflammatory bowel disease, as previously reported¹⁰. Furthermore, consistent with regulation of CD1d expression by HSP110 (ref. 16), oxazolone colitis was associated with induction of CD1d expression in wild-type but not *Mtpp*^{ΔIEC} IECs (Extended Data Fig. 3c, right). Together, these data indicate that IEC CD1d and MTP are required for epithelial HSP110 upregulation during oxazolone-induced colitis and that alterations in this pathway are observed in human ulcerative colitis.

To understand better the functional consequences of epithelial HSP110 deficiency observed in *Mtpp*^{ΔIEC} and *Cd1d1*^{ΔIEC} mice, bone marrow chimaeras of HSP110-knockout (HSP110-KO) or wild-type mice receiving wild-type bone marrow were studied. Compared with wild-type recipients, HSP110-knockout recipients of wild-type bone marrow exhibited increased mortality and weight loss (Fig. 2d), microscopic injury (Fig. 2e) and increased IL-13 and IL-1β secretion (Fig. 2f) in the oxazolone model. Morbidity and mortality in HSP110-knockout recipients of wild-type bone marrow reflected that observed for conventional HSP110-knockout mice, suggesting critical, protective effects of HSP110 in the radioresistant compartment, including IECs (Extended Data Fig. 6c). Antibody-mediated neutralization of CD1d reduced the severity of colitis in HSP110-knockout bone marrow chimaeras, in line with pathogenic CD1d-dependent signals originating from non-epithelial lamina propria cells (Fig. 2d, f). Together, these studies indicate

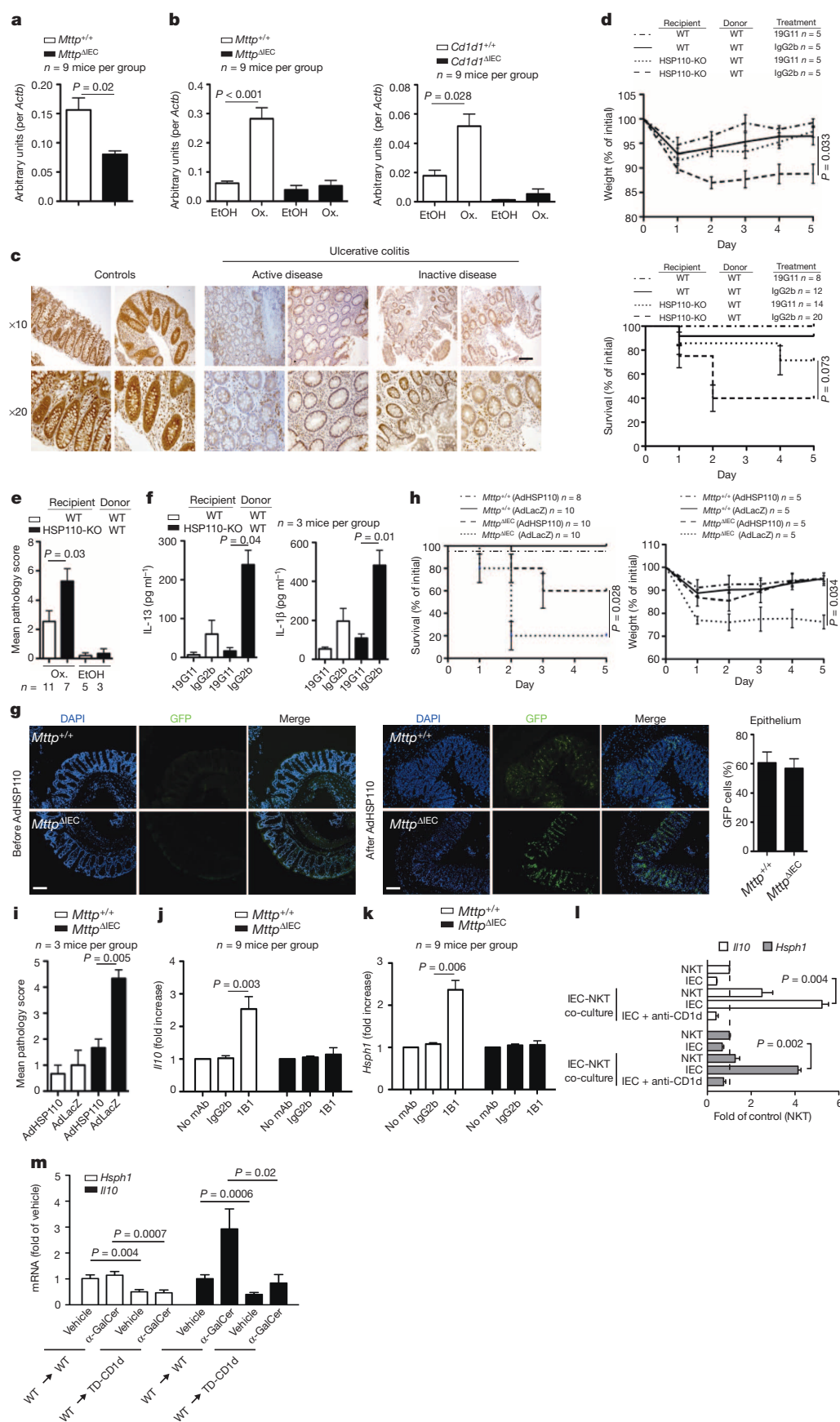


Figure 2 | HSP110 elicits protective effects downstream of intestinal epithelial CD1d. **a, b,** *Hsph1* expression (qPCR) in purified IECs of the indicated mouse strains at baseline (**a**) or on day 5 after rectal oxazolone (Ox.) or ethanol (EtOH) (**b**). **c,** HSP110 immunohistochemistry in colon of human ulcerative colitis and controls. **d–i,** Survival (**d**, bottom, **h**, left), body weight (**d**, top, **h**, right), histopathology (**e**, **i**), and cytokine secretion in colon explants (ELISA) (**f**) in the oxazolone model. **d, f,** Wild-type (WT) and HSP110-KO recipients of wild-type bone marrow received 19G11 anti-CD1d or isotype control antibody. **g–i,** Mice received adenoviruses expressing HSP110 (AdHSP110) or β -galactosidase (AdLacZ). **g,** Immunofluorescence of colonic GFP after AdHSP110-IRES2-eGFP administration. DAPI, 4',6-diamidino-2-phenylindole. **j–l,** MODE-K *Hsph1* and *Il10* (qPCR) after antibody-mediated CD1d crosslinking (1B1) (**j–k**) or α -GalCer presentation to iNKT cells (24.7 hybridoma) in the presence/absence of neutralizing CD1d antibodies (19G11) (**l**). **mAb,** monoclonal antibody. **m,** *Hsph1* and *Il10* (qPCR) in purified small intestinal IECs from wild-type and TD-CD1d recipients of wild-type bone marrow after injection with vehicle or α -GalCer (6 mice per group). **c, g,** Scale bars, 50 μ m. Results representative of two or three independent experiments are shown. Means \pm s.e.m. of the indicated number of mice or triplicate cultures (**l**) are indicated. Student's *t*-test or log-rank test (survival) were applied.

that reduced HSP110 expression in the absence of functional epithelial CD1d may contribute to severe oxazolone colitis. Consistent with this hypothesis, reconstitution of HSP110 expression by adenoviruses targeting

IECs (Fig. 2g) ameliorated oxazolone-induced colitis in *Mtpp*^{ΔIEC} mice, as shown by decreased mortality (Fig. 2h, left), weight loss (Fig. 2h, right) and microscopic injury (Fig. 2i). It is noteworthy that adenoviruses also

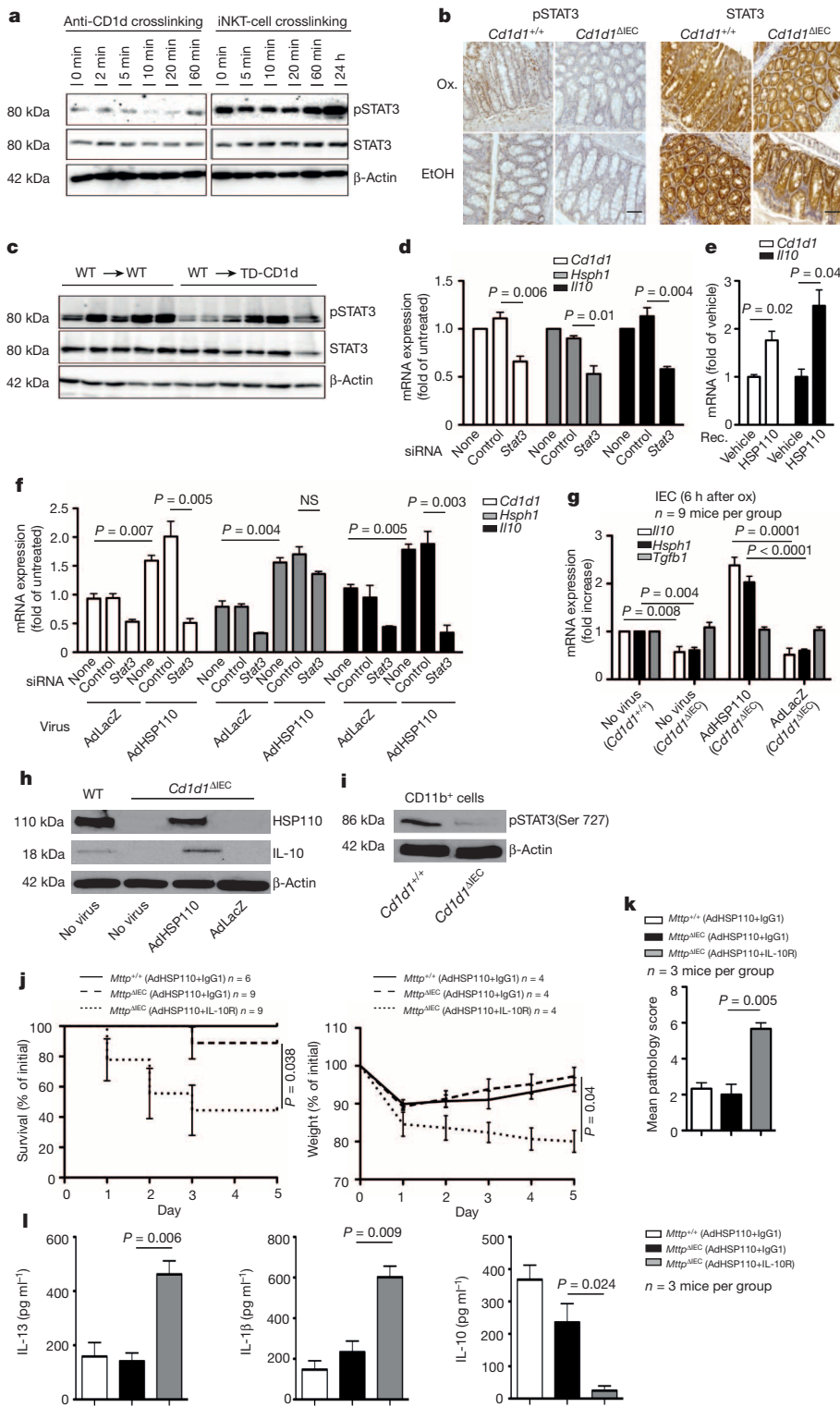


Figure 3 | Epithelial CD1d induces STAT3-dependent expression of IL-10 and HSP110. **a**, MODE-K pSTAT3 after CD1d-crosslinking (1B1 monoclonal antibody clone) or iNKT (24.7 hybridoma) co-culture. **b**, Colonic pSTAT3 immunohistochemistry 5 days after rectal oxazolone (Ox.) or ethanol (EtOH). Scale bar, 50 μ m. **c**, pSTAT3 in purified small intestinal IECs from the indicated bone marrow chimaeras (IECs from one mouse per lane). **d–f**, *Cd1d*, *Hsp1* and *Il10* (qPCR) in MODE-K cells after treatment with recombinant HSP110 (**e**) or *Stat3*/control siRNA knockdown (**d**, **f**) followed by viral transduction at a multiplicity of infection (MOI) of 5 (**f**). **g**, *Il10*, *Hsp1* and *Tgfb* in purified IECs (qPCR) 6 h after rectal oxazolone in the presence or absence of viral infection. The fold change compared to oxazolone-treated *Cd1d*^{+/+} mice without virus is indicated. **h**, Western blot of IECs 24 h after rectal oxazolone. **i**, Western blot of purified CD11b⁺ lamina propria cells 5 days after rectal oxazolone. **j–l**, Survival (**j**, left), body weight (**j**, right), histopathology (**k**) and cytokine secretion by colon explants (ELISA; **l**) of the indicated oxazolone-challenged mice with or without viral infection and antibody treatment as indicated. Results representative of three independent experiments are shown. Means \pm s.e.m. of the indicated number of mice or triplicate cultures (**d–f**) are indicated. Student's *t*-test or log-rank test (survival) were applied. NS, not significant.

targeted lamina propria cells (Fig. 2g), which may have contributed to protection from intestinal inflammation.

As mice with an IEC-specific deficiency in MTP and CD1d exhibited impaired epithelial expression of HSP110, we tested whether engagement of epithelial CD1d supports HSP110 expression. To this end, we investigated HSP110 in purified (Extended Data Fig. 7) colonic IECs after antibody-mediated CD1d crosslinking, a treatment previously shown to induce production of barrier-protective IEC-derived IL-10 (ref. 9). CD1d crosslinking led to increased expression of *Il10* (Fig. 2j) and *Hsp1* (Fig. 2k) in IECs from wild-type but not *Mtpp*^{ΔIEC} mice. Similar observations

were made upon *in vitro* co-culture of NKT cells with a mouse IEC line (Fig. 2l). Previous studies in a cultured human IEC line suggested that epithelial IL-10 production upon CD1d crosslinking is elicited through retrograde signalling dependent on the cytoplasmic tail of CD1d⁹. We therefore investigated whether bone marrow chimaeras with selective deletion of the cytoplasmic CD1d tail (hereafter referred to as tail-deleted (TD)-CD1d) in radioresistant cells exhibit impaired expression of *Il10* and *Hsp1* in purified IECs. Indeed, IECs obtained from these mice showed impaired expression of *Il10* and *Hsp1* under constitutive conditions and reduced expression after administration of the CD1d-binding,

iNKT-cell-activating lipid α -galactosylceramide (α -GalCer) (Fig. 2m). These data demonstrate that engagement of epithelial CD1d induces expression of HSP110 and IL-10, presumably through retrograde CD1d signalling. Additional effects of the CD1d cytoplasmic tail, such as regulation of intracellular CD1d trafficking¹⁷ and potential effects on NKT cell homing⁵, may further contribute to protection from intestinal inflammation.

As the *Il10*, *Hsph1* and *Cd1d1* genes contain STAT sites in their promoters (ref. 18 and T.O. *et al.*, unpublished observations), and as *Hsph1* and *Cd1d1* exhibit reduced expression in *Stat3*-deficient IECs (Gene Expression Omnibus accession number GSE15955)¹⁹, we reasoned that STAT3 may contribute to *Il10* and *Hsph1* expression downstream of CD1d. Indeed, epithelial STAT3 became phosphorylated in response to CD1d engagement by antibody crosslinking (Fig. 3a, left) and iNKT-cell co-culture (Fig. 3a, right, and Extended Data Fig. 8a). Moreover, impaired epithelial STAT3 phosphorylation was observed in *Cd1d1*^{ΔIEC} mice (Fig. 3b) and TD-CD1d recipients of wild-type bone marrow (Fig. 3c). Furthermore, short interfering RNA (siRNA)-mediated knockdown of *Stat3* in the IEC line MODE-K (Extended Data Fig. 8b) led to reduced expression of *Cd1d1*, *Hsph1* and *Il10* (Fig. 3d). Together, these data demonstrate that STAT3 acts downstream of epithelial CD1d to induce expression of *Il10*, *Hsph1* and *Cd1d1*.

IL-10 and HSP110 do not only exhibit STAT3-dependent expression but also signal via STAT3 (refs 18, 20) and may thus further contribute to sustained activation of epithelial STAT3 and STAT3-dependent expression of *Cd1d1*, *Il10* and *Hsph1*. Consistent with this concept, both recombinant and adenoviral HSP110 induced epithelial expression of *Il10* and *Cd1d1* in MODE-K cells, which was abrogated upon siRNA-mediated knockdown of *Stat3* (Fig. 3e, f). Similar findings were made *in vivo*, as adenoviruses expressing HSP110, but not LacZ, restored epithelial expression of IL-10 in *Cd1d1*^{ΔIEC} mice in the oxazolone model

(Fig. 3g, h). These data demonstrate that HSP110 acts in a STAT3-dependent manner to support the expression of IL-10 and CD1d by IECs.

Given HSP110-dependent regulation of epithelial IL-10 expression, we investigated whether the protective effects of HSP110 are dependent on IL-10. CD11b⁺ lamina propria cells from *Cd1d1*^{ΔIEC} mice showed decreased phosphorylated (p)STAT3 expression in oxazolone colitis, consistent with decreased IL-10 production by IECs and IL-10 receptor (IL-10R) signalling by lamina propria cells (Fig. 3i). Moreover, antibody-mediated IL-10R blocking prevented adenoviral HSP110-dependent amelioration of colitis (Fig. 3j–l). To investigate further the role of epithelial IL-10, we generated *Il10*^{−/−} bone marrow chimaeras. *Il10* deletion in the radioresistant compartment resulted in increased mortality (Fig. 4a) and colonic IL-13 production (Fig. 4b) in the oxazolone model. Moreover, a similar trend was observed in mice with *Il10* deletion in the radiosensitive compartment (Fig. 4a, b). These data thus reveal that, in addition to the established role of bone-marrow-derived cells in IL-10-mediated control of intestinal inflammation^{21,22}, IL-10 derived from radioresistant cells provides an additional, non-redundant pathway required for regulation of intestinal inflammation. To address whether protective IL-10 originates from the intestinal epithelium, we investigated oxazolone colitis in mice with inducible, epithelial-specific deletion of *Il10* (Villin-CreER^{T2} × *Il10*^{fl/fl}; hereafter referred to as *Il10*^{ΔIEC}). Similar to mice with IEC-specific deletion of *Mttrp* and *Cd1d1* and mice with deletion of *Hsph1* and *Il10* in the radioresistant compartment, *Il10*^{ΔIEC} mice exhibited increased mortality (Fig. 4c), weight loss (Fig. 4d) and histopathological inflammation (Fig. 4e) in the oxazolone model. Furthermore, in accordance with STAT3-mediated signalling by IL-10 (ref. 18) and STAT3-dependent expression of *Hsph1* and *Cd1d1* (Fig. 3d), IEC-specific *Il10* deletion was associated with reduced epithelial expression of *Hsph1* and a trend towards a reduction in *Cd1d1* expression in the oxazolone model (Fig. 4f). These results demonstrate that IEC-derived

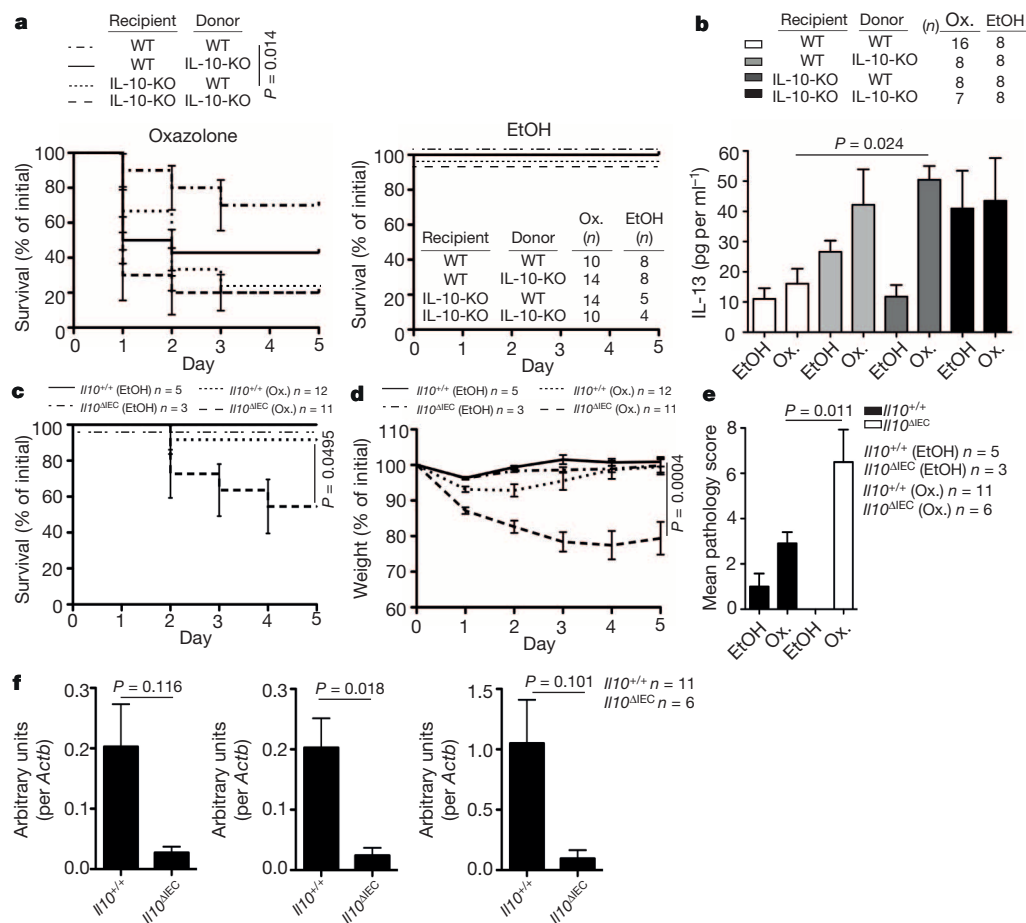


Figure 4 | Epithelial IL-10 is critical for control of intestinal inflammation.

a–f, Survival (a, c), IL-13 production in colon explants on day 1 after rectal oxazolone (Ox.) or ethanol (EtOH) (b), body weight (d), histopathology (e), and *Il10* (left), *Hsph1* (middle) and *Cd1d1* (right) expression (qPCR; f) in colonic IECs 5 days after rectal oxazolone challenge in the indicated mouse strains. Results representative of three (a, b) or two (c–f) independent experiments are shown. Means \pm s.e.m. of the indicated number of mice are shown. Student's *t*-test or log-rank test (survival) were applied.

IL-10, downstream of epithelial CD1d, HSP110 and STAT3 is critical for regulation of intestinal inflammation.

NKT cells are central mediators of intestinal inflammation^{3,4}. In studies aiming to identify the cellular origin of pathogenic lipid antigen presentation in oxazolone colitis, we made the unanticipated observation that intestinal epithelial CD1d protects from intestinal inflammation in a manner dependent on STAT3, IL-10 and HSP110, whereas pathogenic NKT-cell activation emanates from bone-marrow-derived cells in a CD1d-restricted fashion (Extended Data Fig. 1a). As further outlined in the Supplementary Discussion and Extended Data Fig. 9, such opposing roles of CD1d-restricted presentation by bone-marrow-derived and epithelial cells are at least partially dependent on the expression of co-stimulatory molecules such as CD40. Our findings highlight the role of the intestinal epithelium as a key regulator of mucosal immunity and underscore the important role of intestinal epithelial CD1d and IL-10 in determining the host's response to environmental stimuli, which activate CD1d-restricted NKT cells⁵.

METHODS SUMMARY

Mice were housed in a specific pathogen-free (SPF) barrier facility. HSP110-KO (ref. 23), *Cd1d1*^{-/-} *Cd1d2*^{-/-} (ref. 24), TD-CD1d (ref. 17), *Il10*^{-/-} (ref. 25), Villin-CreERT² (ref. 14), *Mttr*^{fl/fl} (ref. 15), *Il10*^{fl/fl} (ref. 26), and *Rag1*^{-/-} (ref. 27) mice have been described. For *Cd1d1*^{ΔIEC} mice, please refer to Methods. To activate Cre recombinase, 4–5-week-old Villin-CreERT² mice were orally gavaged daily with 1 mg tamoxifen for a total of 5 days. For bone marrow chimaeras, recipients received total body irradiation of 1,100 rad in two separate doses 4 h apart. The next day 0.5 × 10⁶ bone marrow cells were delivered intravenously. Oxazolone colitis was induced as described previously⁴ and as outlined in Methods. Where indicated, mice received 1 mg of neutralizing antibodies against IL-10R (R&D), CD1d (19G11, BioXcell) or isotype control intraperitoneally (i.p.), or 5 × 10⁸ adenoviruses in PBS i.p., each before skin sensitization and before and 2 days after rectal oxazolone challenge. For a schematic overview of procedures please refer to Extended Data Fig. 1b.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2012; accepted 17 February 2014.

Published online 6 April 2014.

- Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
- Saenz, S. A., Taylor, B. C. & Artis, D. Welcome to the neighborhood: epithelial cell-derived cytokines license innate and adaptive immune responses at mucosal sites. *Immunol. Rev.* **226**, 172–190 (2008).
- Fuss, I. J. et al. Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J. Clin. Invest.* **113**, 1490–1497 (2004).
- Heller, F., Fuss, I. J., Nieuwenhuis, E. E., Blumberg, R. S. & Strober, W. Oxazolone colitis, a Th2 colitis model resembling ulcerative colitis, is mediated by IL-13-producing NK-T cells. *Immunity* **17**, 629–638 (2002).
- Olszak, T. et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012).
- Wingender, G. & Kronenberg, M. Role of NKT cells in the digestive system. IV. The role of canonical natural killer T cells in mucosal immunity and inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* **294**, G1–G8 (2008).
- Heller, F. et al. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* **129**, 550–564 (2005).
- Liao, C. M. et al. Dysregulation of CD1d-restricted type II natural killer T cells leads to spontaneous development of colitis in mice. *Gastroenterology* **142**, 326–334 (2012).
- Colgan, S. P., Hershberg, R. M., Furuta, G. T. & Blumberg, R. S. Ligation of intestinal epithelial CD1d induces bioactive IL-10: critical role of the cytoplasmic tail in autocrine signaling. *Proc. Natl Acad. Sci. USA* **96**, 13938–13943 (1999).
- Perera, L. et al. Expression of nonclassical class I molecules by intestinal epithelial cells. *Inflamm. Bowel Dis.* **13**, 298–307 (2007).
- Blumberg, R. S. et al. Expression of a nonpolymorphic MHC class I-like molecule, CD1d, by human intestinal epithelial cells. *J. Immunol.* **147**, 2518–2524 (1991).
- Brozovic, S. et al. CD1d function is regulated by microsomal triglyceride transfer protein. *Nature Med.* **10**, 535–539 (2004).
- Zeissig, S. et al. Primary deficiency of microsomal triglyceride transfer protein in human abetalipoproteinemia is associated with loss of CD1 function. *J. Clin. Invest.* **120**, 2889–2899 (2010).
- El Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
- Raabe, M. et al. Analysis of the role of microsomal triglyceride transfer protein in the liver of tissue-specific knockout mice. *J. Clin. Invest.* **103**, 1287–1298 (1999).
- Colgan, S. P. et al. Intestinal heat shock protein 110 regulates expression of CD1d on intestinal epithelial cells. *J. Clin. Invest.* **112**, 745–754 (2003).
- Chiu, Y. H. et al. Multiple defects in antigen presentation and T cell development by mice expressing cytoplasmic tail-truncated CD1d. *Nature Immunol.* **3**, 55–60 (2002).
- Saraiva, M. & O'Garra, A. The regulation of IL-10 production by immune cells. *Nature Rev. Immunol.* **10**, 170–181 (2010).
- Pickert, G. et al. STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* **206**, 1465–1472 (2009).
- Yamagishi, N., Fujii, H., Saito, Y. & Hatayama, T. Hsp105 β upregulates *hsp70* gene expression through signal transducer and activator of transcription-3. *FEBS J.* **276**, 5870–5880 (2009).
- Chaudhry, A. et al. Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity* **34**, 566–578 (2011).
- Rubtsov, Y. P. et al. Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28**, 546–558 (2008).
- Nakamura, J. et al. Targeted disruption of Hsp110/105 gene protects against ischemic stress. *Stroke* **39**, 2853–2859 (2008).
- Smiley, S. T., Kaplan, M. H. & Grusby, M. J. Immunoglobulin E production in the absence of interleukin-4-secreting CD1-dependent cells. *Science* **275**, 977–979 (1997).
- Kühn, R., Lohler, J., Rennick, D., Rajewsky, K. & Müller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
- Roers, A. et al. T cell-specific inactivation of the interleukin 10 gene in mice results in enhanced T cell responses but normal innate responses to lipopolysaccharide or skin irritation. *J. Exp. Med.* **200**, 1289–1297 (2004).
- Mombaerts, P. et al. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank H.-C. Hung for technical assistance with microinjection, Y. Xie for performing osmium staining, A. Bedynek and M. Friedrich for performing immunohistochemistry of the human biopsies, F. A. Zhu for assistance with antigen presentation assays, D. Shouval, M. Sablon and D. Perez for animal care and husbandry, K. Tashiro for technical assistance with adenovirus preparation, V. M. Thiele for technical assistance, J. Cusick for help with manuscript preparation, and S. E. Plevy for discussions and reagents. This work was supported by: National Institutes of Health (NIH) (grants DK044319, DK051362, DK053056, DK088199) and the Harvard Digestive Diseases Center (DK0034854) (R.S.B.); the European Research Council (ERC Starting Grant agreement no. 336528), the Deutsche Forschungsgemeinschaft (DFG) (ZE 814/4-1, ZE 814/5-1, ZE 814/6-1), the Crohn's and Colitis Foundation of America (Postdoctoral Fellowship Award), the European Commission (Marie Curie International Reintegration Grant no. 256363) and the DFG Excellence Cluster "Inflammation at Interfaces" (S.Z.); the DFG (OL 324/1-1) (T.O.); HL38180, DK56260, Washington University DDRCC P30DK52574 (morphology core) (N.O.D.); HDDC Pilot and Feasibility Grant (K.B.); NCI P30CA013696 (C.-S.L.); the DFG (BR 1912/6-1) and the Else Kroener-Fresenius-Stiftung (Else Kroener-Exzellenzstipendium 2010_EKES.32) (S.B.); Grant-in-Aid for Challenging Exploratory Research 24659823 from Japan Society for Promotion of Science (K.W.); the ERC under the European Community's Seventh Framework Programme (FP7/2007-2013/ERC Grant agreement no. 260961), the National Institute for Health Research Cambridge Biomedical Research Centre, the Austrian Science Fund and Ministry of Science P21530-B18 and START Y446-B18, Innsbruck Medical University (MFI 2007-407) and the Addenbrooke's Charitable Trust, CiCRA (A.K.); the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement SysmedIBD (no. 305564) (W.M., S.S.); the NIH (grants HL59561, DK034854, AI50950), the Helmsley Charitable Trust and the Wolpew Family Chair in IBD Treatment and Research (S.B.S.). PBS57-loaded and unloaded mouse CD1d tetramer was obtained through the NIH Tetramer Facility. The authors thank M. A. Exley and S. P. Colgan for discussions.

Author Contributions T.O., J.F.N., C.M.D. and K.B. performed *in vitro* and *in vivo* experiments and analysed the results; N.O.D. performed osmium tetroxide staining; J.G. obtained and scored histopathologies; C.-S.L. generated *Cd1d1*^{fl/fl} mice; C.J. contributed to the analysis of CD1d^{ΔIEC} mice; S.B. and K.S. contributed to the immunohistochemical analysis of ulcerative colitis patients; K.W., K. Katayama, A.N. and H.M. generated adenoviruses; K. Kawasaki and K.N. provided HSP110-KO mice; W.M. and S.B.S. provided and participated in the analysis of the *Il10*^{ΔIEC} mice; S.S. contributed to the coordination of experimental studies; A.K. contributed to *Mttr*^{ΔIEC} studies and to the analysis of microarray data; R.S.B. and S.Z. designed the study, coordinated the experimental work and wrote the manuscript with input from co-authors. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.B. (rblumberg@partners.org) or S.Z. (szeissig@1med.uni-kiel.de).

METHODS

Mice. Mice (C57BL/6J unless indicated otherwise) were housed in a specific pathogen-free (SPF) barrier facility. *Cd1d1*^{-/-} *Cd1d2*^{-/-} (ref. 24), TD-CD1d (ref. 17), *Il10*^{-/-} (ref. 25), Villin-CreER^{T2} (ref. 14), *Il10*^{fl/fl} (ref. 26), *Rag1*^{-/-} (ref. 27), HSP110-KO (ref. 23) (C57BL/6N) and *Mtpp*^{fl/fl} (ref. 15) (C57BL/6J/129) mice have been described. To generate the floxed/neo *Cd1d1* allele, a LoxP (L83) site at 10310 and a FNFL (Frt-Neo-Frt-LoxP) cassette at 12190 was inserted to flank exon 2, 3 and 4 (about 1.9 kb) of the *Cd1d1* gene. A gene-targeting vector was constructed by retrieving one 5 kb long homology arm (5' to L83), one 1.9 kb sequence containing exon 2/3/4, FNFL cassette, and one 2 kb short homology arm (end of FNFL to 3'). The FNFL cassette conferred G418 resistance during gene targeting in PTL1 (129B6 hybrid) embryonic stem (ES) cells. The *Cd1d1* allele was PCR amplified and sequenced to confirm the targeted C57BL/6J allele based on the C57BL/6J-specific mutation in *Cd1d2* allele²⁸. Three targeted ES cells with targeted C57BL/6J allele were injected into C57BL/6J blastocysts to generate chimaeras. Male chimaeras were bred to bACTFpe females or EIIa-Cre females to transmit the floxed *Cd1d1* allele (*Cd1d1*^{fl/+}) (with neo cassette removed by Ppe recombinase) and *Cd1d1*-null allele (*Cd1d1*^{+/-}) (with exon 2/3/4 and neo cassette removed by Cre recombinase) through the germ line. Mice carrying floxed *Cd1d1* alleles were crossed to Villin-CreER^{T2} mice (ref. 14) expressing Cre recombinase.

To activate Cre recombinase, 4–5-week-old Villin-CreER^{T2} mice were orally gavaged daily with 1 mg tamoxifen for a total of 5 days. For bone marrow chimaeras, recipients received total body irradiation of 1,100 rad in two separate doses 4 h apart. The next day, 0.5×10^6 bone marrow cells were delivered intravenously. For α -GalCer studies, mice were injected with 2 μ g of α -GalCer (KRN7000, Avanti Polar Lipids) intraperitoneally (i.p.) 8 weeks after bone marrow reconstitution and RNA was isolated from Percoll-purified IECs (see later) 12 h after injection. For oxazolone colitis, mice were pre-sensitized by abdominal skin application of 3% (w/v) oxazolone in 100% ethanol, followed by rectal administration of 1% oxazolone in 50% ethanol after 5 days. Where indicated, mice received 1 mg of neutralizing antibodies against IL-10R (R&D), CD1d (19G11, BioXcell) or isotype control i.p., or 5×10^8 adenoviruses in PBS i.p., each before skin sensitization and before and 2 days after rectal oxazolone challenge. Body weight, rectal bleeding and stool consistency were analysed on a daily basis. Tissues were examined for evidence of colitis by five established criteria for colitis activity in a blinded fashion by a pathologist (J.G.): hypervascularization, presence of mononuclear cells, epithelial hyperplasia, epithelial injury, and presence of granulocytes. For a schematic overview of procedures please refer to Extended Data Fig. 1b.

Animal studies were conducted in a gender- and age-matched manner using littermates for each experiment. Both male and female mice were used and were 8 weeks of age at the time of experiments (11–16 weeks of age for bone marrow chimaera studies). The number of animals used per group was based on previous experimental results and observed variability. Histopathology analysis was performed in a blinded manner. For all other *in vitro* and *in vivo* analyses, investigators were not blinded to treatment allocation. Animal studies were performed in compliance with ethical regulations and were approved by the ethics committees of Harvard Medical School and the Christian Albrechts University.

Patients. Human studies were approved by the ethics committee of the Ludwig Maximilians University. All subjects provided written informed consent. Immunohistochemical staining was performed on intestinal biopsies from ten healthy controls (patients undergoing screening colonoscopy without detection of malignancy), four patients with active ulcerative colitis, and five patients with inactive ulcerative colitis.

Isolation of IECs and lymphocytes, flow cytometry, and sorting. To isolate IECs, the intestines were opened, washed in cold PBS, cut into pieces, and transferred into tubes containing HBSS without calcium and magnesium. After the addition of dithiothreitol (DTT) at 1 mM (Sigma), intestines were shaken at 250 r.p.m. for 10 min at room temperature to remove mucus. After washing in PBS, tissues were incubated in RPMI-1640 with 1 U ml⁻¹ Dispase (Roche) for 30–40 min at 37 °C at 250 r.p.m. After filtering through a 100 μ m strainer and centrifugation for 5 min at 1,500 r.p.m., the pellet was resuspended in 5 ml of 100% Percoll and placed beneath 8 ml of 40% Percoll. After centrifugation for 20 min at 1,500 r.p.m. the top layer was removed, washed three times in PBS and used as IECs.

To isolate intraepithelial (IEL) and lamina propria lymphocytes, large intestines were collected, fat tissue removed, the intestines cut longitudinally and washed in PBS in order to remove faecal content, cut into 30 mm pieces, and shaken in RPMI-1640 containing 20 mM HEPES, 1 mM DTT and 5 mM EDTA for 30 min at 37 °C. To isolate IELs, cells derived from the epithelial compartment were resuspended in 4 ml of 40% Percoll and layered on top of 2 ml of 75% Percoll. After centrifugation for 20 min at 2,000 r.p.m. the middle layer was removed, washed in 2% FBS in PBS and the IELs were obtained. For lamina propria lymphocytes, the non-epithelial compartment was washed in PBS, cut into smaller pieces and incubated in RPMI-1640 containing 5% FBS, 1.5 mg ml⁻¹ collagenase type II and 0.5 mg ml⁻¹ dispase (GIBCO) for 1 h at 37 °C under constant horizontal shaking (250 r.p.m.). The digested

tissues were filtered through a 40 μ m strainer and, after Percoll gradient centrifugation and washing as described earlier, were used as lamina propria lymphocytes in flow cytometry. Where indicated, CD11b⁺ lamina propria cells were purified using magnetic beads according to the manufacturer's instructions (Miltenyi Biotec). Liver lymphocytes were prepared as described before²⁹.

Primary iNKT cells were sorted by double staining with PBS57-loaded CD1d tetramers and CD3 ϵ using a BD FACSAria II SORP UV sorter. RNA samples were prepared using an RNeasy Micro Kit and cDNAs were synthesized using the Omniscript RT Kit (Qiagen).

PBS57-loaded or -unloaded CD1d tetramers were obtained from the NIH Tetramer Core Facility. Flow cytometry antibodies were obtained from eBiosciences. Flow cytometry was performed using a MACSQuant (Miltenyi Biotec) and a BD FACSVerser (BD Biosciences) and data were analysed with FlowJo software (TreeStar).

Antigen presentation assays. Freshly isolated IECs, MODE-K cells³⁰, and CD11c-magnetic-bead-purified (Miltenyi Biotec) splenic dendritic cells were incubated with 100 ng ml⁻¹ α -GalCer for 4 h, washed three times with PBS, and aliquoted into 96-well plates at 1×10^5 cells per well. NKT-cell hybridomas (24.7, DN32.D3 and 14S.6 (ref. 31)) were added at a 1:1 ratio. Where indicated, soluble, stimulatory CD28 antibody was added (1 μ g ml⁻¹; clone 37.51; eBioscience). In addition, where indicated, cells were transfected with expression plasmids or siRNA 24 h and 72 h, respectively, before co-culture with NKT cells. Mouse IL-2, IL-10, IL-12p70, IL-13 and IL-1 β production were assessed by ELISA (OptEIA; BD Biosciences) in supernatants after 24 h of co-culture.

Cell stimulation by crosslinking. CD1d crosslinking on freshly isolated colonic IECs was achieved as described previously³² using plate-bound anti-mCD1d antibody (1B1, BD). In the case of MODE-K cells, surface CD1d was crosslinked by addition of soluble anti-mCD1d antibody (1B1, 10 μ g ml⁻¹, 60 min) followed by washing and addition of secondary goat anti-rat IgG (Biolegend, 10 μ g ml⁻¹) for the indicated time as well as subsequent cell lysis (see later). MODE-K STAT3 phosphorylation in response to iNKT-cell engagement was investigated after co-culture with the iNKT-cell hybridoma 24.7 for the indicated time, removal of non-adherent iNKT cells by thorough washing, and cell lysis. Where indicated, MODE-K cells were pre-treated with 19G11 anti-CD1d antibody for blocking of CD1d (10 μ g ml⁻¹, BioXcell).

Adenoviruses. Type 5, E1/E3-deleted adenoviruses expressing murine HSP110 (AdHSP110) or *Escherichia coli* β -galactosidase (AdLacZ) under the control of the cytomegalovirus promoter, were constructed according to the *in vitro* ligation method^{33,34}. Briefly, murine *Hsp11* (NM-013559) was isolated and amplified by PCR from mouse intestinal cDNA using the following primers, cloned into NheI and EcoRI sites of pIRES2-eGFP (Clontech), sequenced, and the resulting plasmid was called pHSP110-IRES2-eGFP. Forward primer: ATAAAGCTAGCGCCACCatgtc ggtggtggcgtag, containing an NheI site and a Kozak like sequence; reverse primer: AATTGAATTCCTAGtcaggtccatgttgacagacg, containing an EcoRI site (the sequence complementary to *Hsp11* cDNA is shown in lower case). An NheI-HSP110-KpnI or NheI-HSP110-IRES2-eGFP-NotI fragment from pHSP110-IRES2-eGFP was transferred into pHMCMV6 (pHMCMV6-HSP110 and pHMCMV6-HSP110-IRES2-eGFP, respectively). Subsequently, pAdHM15RGD-CMV6-HSP110 and pAdHM15RGD-CMV6-HSP110-IRES2-eGFP were constructed by ligating I-CeuI/PI-SceI-digested pAdHM15-RGD and pHMCMV6-HSP110-IRES2-eGFP or pHMCMV6-HSP110. To prepare the virus, a PacI fragment of pAdHM15-RGD-CMV-HSP110 or pAdHM15-RGD-CMV-HSP110 was transfected into 293 cells. The adenoviruses were purified by two rounds of caesium chloride density centrifugation and dialysed against Tris-HCl (pH 7.5). The concentrations of plaque-forming units (p.f.u.) of individual stocks were determined from the tissue culture infectious dose 50 titre.

Colon organ culture. Standardized segments (1 cm \times 1 cm) of the transverse colon were washed in cold PBS supplemented with penicillin and streptomycin (GIBCO) and cultured in 24-well flat-bottom culture plates (Falcon) in RPMI 1640 media (GIBCO) for 24 h at 37 °C. Supernatants were analysed for cytokines IL-10, IL-13 and IL-1 β by ELISA (all BD).

Western blotting. Protein extraction and western blotting were performed as described previously³⁵. The following antibodies were used: anti-IL10 (R&D), anti-HSP110 (BD), anti-STAT3 (Cell signaling), anti-pSTAT3 (Cell signaling) and anti- β -actin (Cell Signaling).

Real-time RT-PCR. RNA samples were prepared using an RNeasy Mini Kit and cDNAs were synthesized using the Omniscript RT Kit (Qiagen). Real-time RT-PCR was performed using SYBR Green I Master Mix (Roche) and a CFX96 Real-Time System (Bio-Rad). Values were normalized to β -actin. The following primer sets were used: β -actin, 5'-GATGCTCCCGGGCTGTATT-3' and 5'-GGGGTCACTTCAGGGTCAGGA-3'; *Hsp11*, 5'-CAGGTACAACTGATGGTCAACA-3' and 5'-TGAGGTAAAGTTCAGGTGAAGGG-3'; *Il10*, 5'-GAGAGCTGCAGGGCCTTTTC-3' and 5'-CTCCCTGGTTTCTCTTCCCAAGACC-3'; *Tgfb*, 5'-TGTACGGCAGTGGCTGAACCA-3' and 5'-TGTCACAAGAGCAGTGAGCGCT-3'.

Cd1d1, 5'-GCAGCCAGTACGCTCTTTTC-3' and 5'-ACAGCTTGTTTCTGGCAGGT-3'; *Stat3*, 5'-CCCGTACCTGAAGACCAAGT-3' and 5'-ACACTCCGAGGTCAGATCCA-3'; *Il13*, 5'-AGCATGGTATGGAGTGTGGACCTG-3' and 5'-CAGTTGCTTTGTGTAGCTGAGCAG-3'; *Ifng* 5'-TCAGCAACAGCAAGCGAAAAAGG-3' and 5'-CCACCCGAATCAGCAGCGA-3'.

Transfection, siRNA treatment and recombinant HSP110. Where indicated, MODE-K cells were transfected with mouse CD40 or CD1d (in pSR α -neo), with SignalSilence *Stat3* siRNA II or with SignalSilence Control siRNA (Cell Signaling Technology) using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. *Il10* siRNA was obtained from Invitrogen and transduced by Amaxa Nucleofector technology (Lonza) according to the manufacturer's instructions. Cells were used for antigen presentation assays after 72 h or were incubated with the indicated adenoviruses after 48 h at a MOI of 5 before RNA preparation after another 24 h as described earlier. Where indicated, recombinant HSP110 was added at a final concentration of 6 $\mu\text{g ml}^{-1}$ to MODE-K cells 24 h before RNA extraction.

XBPI splicing assay. The assay was performed as described previously³⁶.

In vivo permeability assay. *In vivo* intestinal permeability was measured using the FITC-labelled dextran (FD-4) method as described previously^{37,38} with modifications. Mice were administered 150 μl of FD-4 (50 mg ml^{-1}) by oral gavage before and 18 h after rectal challenge with oxazolone. Serial dilutions of FD-4 were made to generate a standard curve and serum concentrations of FD-4 were determined using a BioTek FLx800 Fluorescence Microplate Reader (BioTek) with an excitation wavelength of 490 nm and emission wavelength of 530 nm.

Immunohistochemistry and immunofluorescence. A polyclonal rabbit anti-HSP110 antibody (anti-mouse 1:250, anti-human 1:500; Sigma-Aldrich), an anti-STAT3 (anti-mouse 1:900) and anti-pSTAT3 antibody (anti-mouse 1:350), both from Cell Signaling, as well as a rabbit polyclonal anti-GFP antibody conjugated to FITC (1:1,000) (Novus Biologicals) were used. All images were taken by a digital camera on a Nikon Eclipse Ti microscope.

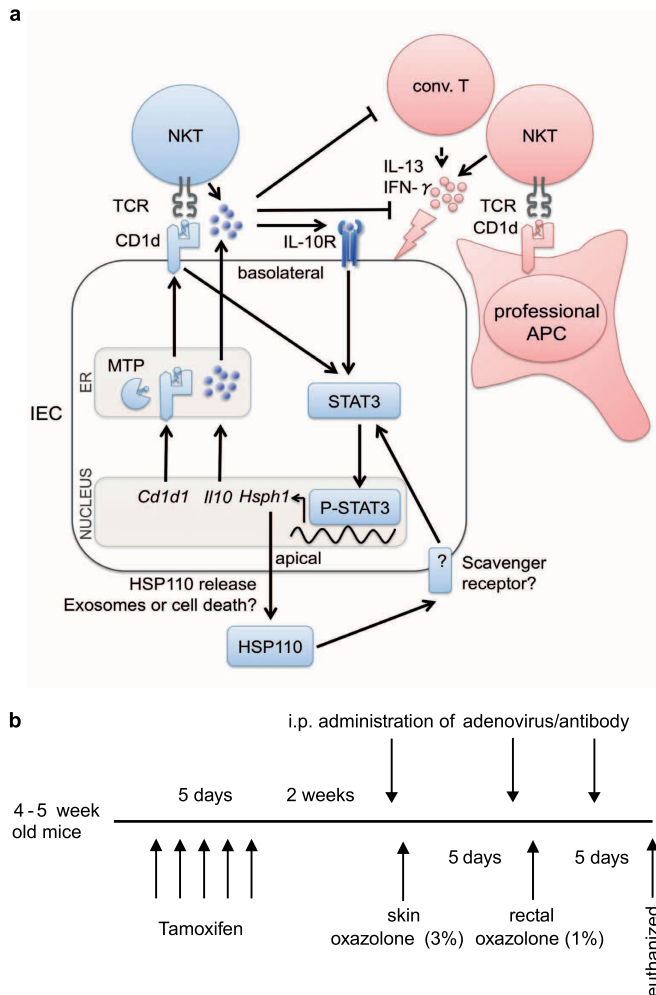
Osmium tetroxide staining of intracellular lipid droplets. Intestinal tissue was fixed in 10% neutral buffered formalin and transferred into 1% osmium tetroxide with periodic shaking. The tissue was rinsed with distilled water and incubated in 0.5% periodic acid, washed and counterstained with haematoxylin and eosin.

Microarray analysis. *Mtpr*^{ΔIEC} and littermate wild-type mice were skin sensitized with oxazolone, and IECs were harvested via colonic epithelial scraping as described previously³⁵, either before administration of rectal oxazolone ($n = 2$ per group) or

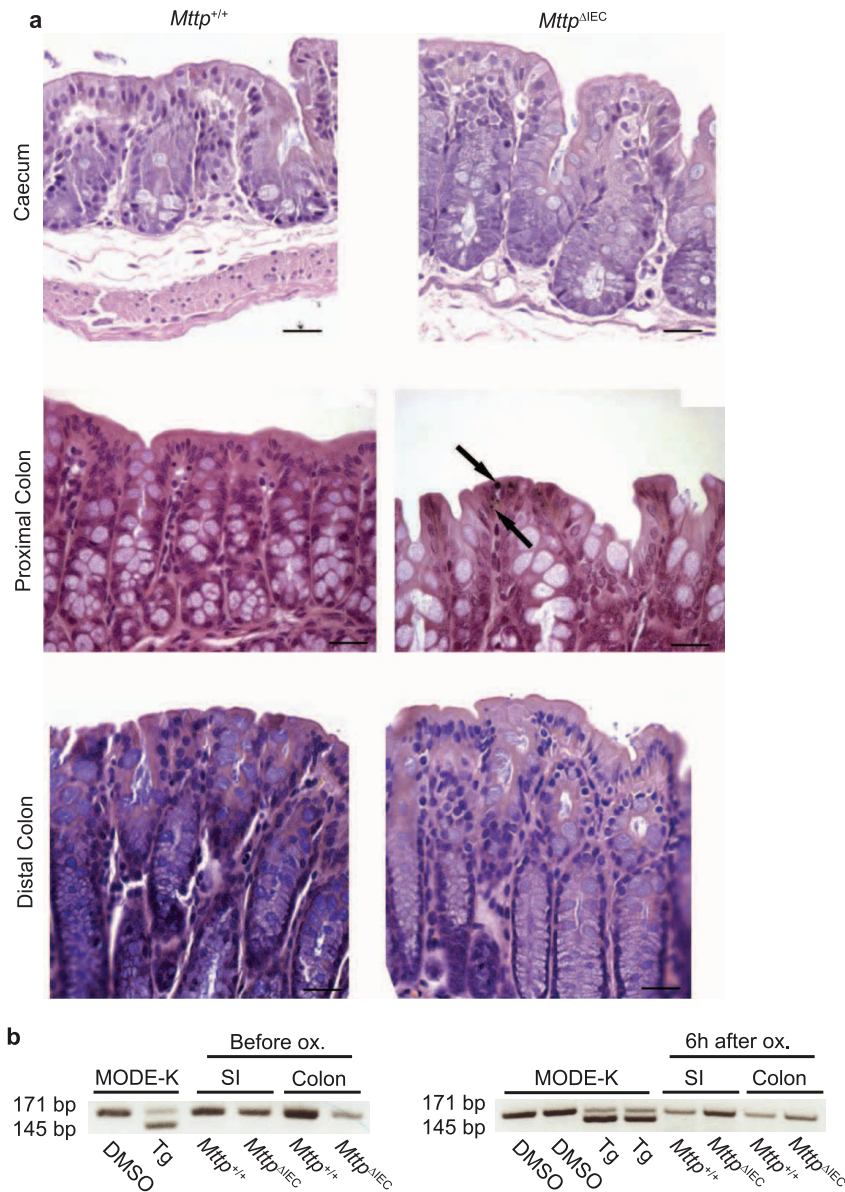
6 h thereafter ($n = 4$ per group). Isolated RNAs were pooled for each condition and subjected to transcriptomic analysis with mouse genome 430 2.0 array (Affymetrix) at the Biopolymers Core Facility (Harvard Medical School). Data analysis was performed with Agilent GeneSpring GX and Affymetrix GCOS software under default parameter settings with GC-RMA normalization.

Statistical analysis. Statistical testing was performed using the unpaired Student's *t*-test unless otherwise indicated. The Mann–Whitney *U*-test was applied when data were demonstrated to not follow a Gaussian distribution, which is indicated in the respective figures legends. Comparisons of mortality were made by analysing Kaplan–Meier survival curves, and the log-rank test was used to assess differences in survival. All *P* values were two-tailed, and statistical significance was assumed at $P < 0.05$.

28. Park, S. H., Roark, J. H. & Bendelac, A. Tissue-specific recognition of mouse CD1 molecules. *J. Immunol.* **160**, 3128–3134 (1998).
29. Zeissig, S. *et al.* Hepatitis B virus-induced lipid alterations contribute to natural killer T cell-dependent protective immunity. *Nature Med.* **18**, 1060–1068 (2012).
30. Vidal, K., Grosjean, I., Revillard, J.-P., Gespach, C. & Kaiserlian, D. Immortalization of mouse intestinal epithelial cells by the SV40-large T gene. Phenotypic and immune characterization of the MODE-K cell line. *J. Immunol. Methods* **166**, 63–73 (1993).
31. Behar, S. M., Podrebarac, T. A., Roy, C. J., Wang, C. R. & Brenner, M. B. Diverse TCRs recognize murine CD1. *J. Immunol.* **162**, 161–167 (1999).
32. Yue, S. C., Shaulov, A., Wang, R., Balk, S. P. & Exley, M. A. CD1d ligation on human monocytes directly signals rapid NF- κ B activation and production of bioactive IL-12. *Proc. Natl Acad. Sci. USA* **102**, 11811–11816 (2005).
33. Katayama, K. *et al.* A novel PPAR γ gene therapy to control inflammation associated with inflammatory bowel disease in a murine model. *Gastroenterology* **124**, 1315–1324 (2003).
34. Mizuguchi, H. & Kay, M. A. A simple method for constructing E1- and E1/E4-deleted recombinant adenoviral vectors. *Hum. Gene Ther.* **10**, 2013–2017 (1999).
35. Chen, Z., Chen, L., Qiao, S. W., Nagaishi, T. & Blumberg, R. S. Carcinoembryonic antigen-related cell adhesion molecule 1 inhibits proximal TCR signaling by targeting ZAP-70. *J. Immunol.* **180**, 6085–6093 (2008).
36. Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).
37. Furuta, G. T. *et al.* Eosinophils alter colonic epithelial barrier function: role for major basic protein. *Am. J. Physiol. Gastrointest. Liver Physiol.* **289**, G890–G897 (2005).
38. Furuta, G. T. *et al.* Hypoxia-inducible factor 1-dependent induction of intestinal trefoil factor protects barrier function during hypoxia. *J. Exp. Med.* **193**, 1027–1034 (2001).

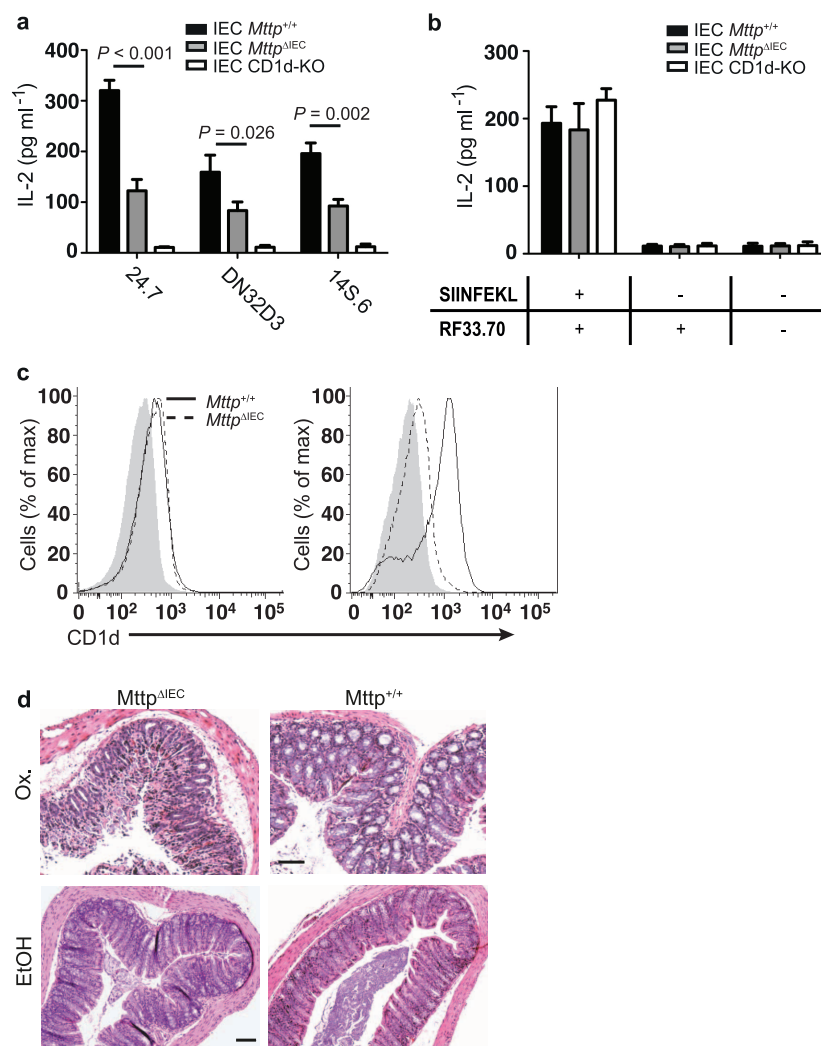


Extended Data Figure 1 | Proposed model of CD1d signalling in polarized intestinal epithelia and overview of experimental procedures. **a**, The proposed model of protective (blue) and pathogenic (red) effects of lipid antigen presentation in intestinal inflammation. Bone-marrow-derived antigen-presenting cells (APCs) contribute to oxazolone colitis in a CD1d- and iNKT-cell-dependent manner. By contrast, engagement of intestinal epithelial CD1d elicits protective functions through cytoplasmic CD1d tail-dependent activation of STAT3, and STAT3-dependent transcription of *Cd1d1*, *Il10* and *Hsph1*. Epithelial IL-10 and HSP110 support this protective self-reinforcing pathway through STAT3-dependent signalling. Interference with any of the elements involved in this regulatory pathway (MTP, CD1d, IL-10 or HSP110) is associated with uncontrolled intestinal inflammation, thus highlighting a critical role of this pathway in the control of intestinal inflammation. Conv., conventional. **b**, Overview of experimental procedures.



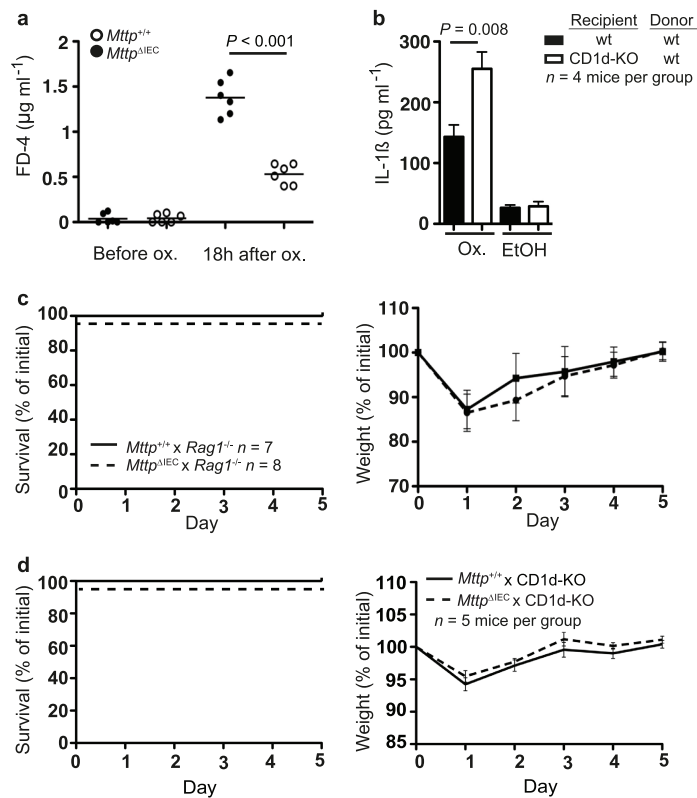
Extended Data Figure 2 | Absence of enterocyte lipid accumulation and ER stress in *Mttr*^{ΔIEC} mice. **a**, Absence of lipid accumulation in *Mttr*^{ΔIEC} mice as shown in representative haematoxylin and eosin and rare osmium tetroxide (arrows) staining in the caecum, proximal, and distal colon. Scale bar, 25 μ m. **b**, *Xbp1* splicing in small intestinal (SI) and colonic IECs of the indicated mice

before (left) ($n = 5$ mice per group) and 6 h after rectal challenge with oxazolone (Ox., right) ($n = 5$ mice per group). MODE-K cells were treated with either thapsigargin (Tg) or vehicle (DMSO) as positive and negative controls, respectively. Results representative of three independent experiments are shown.

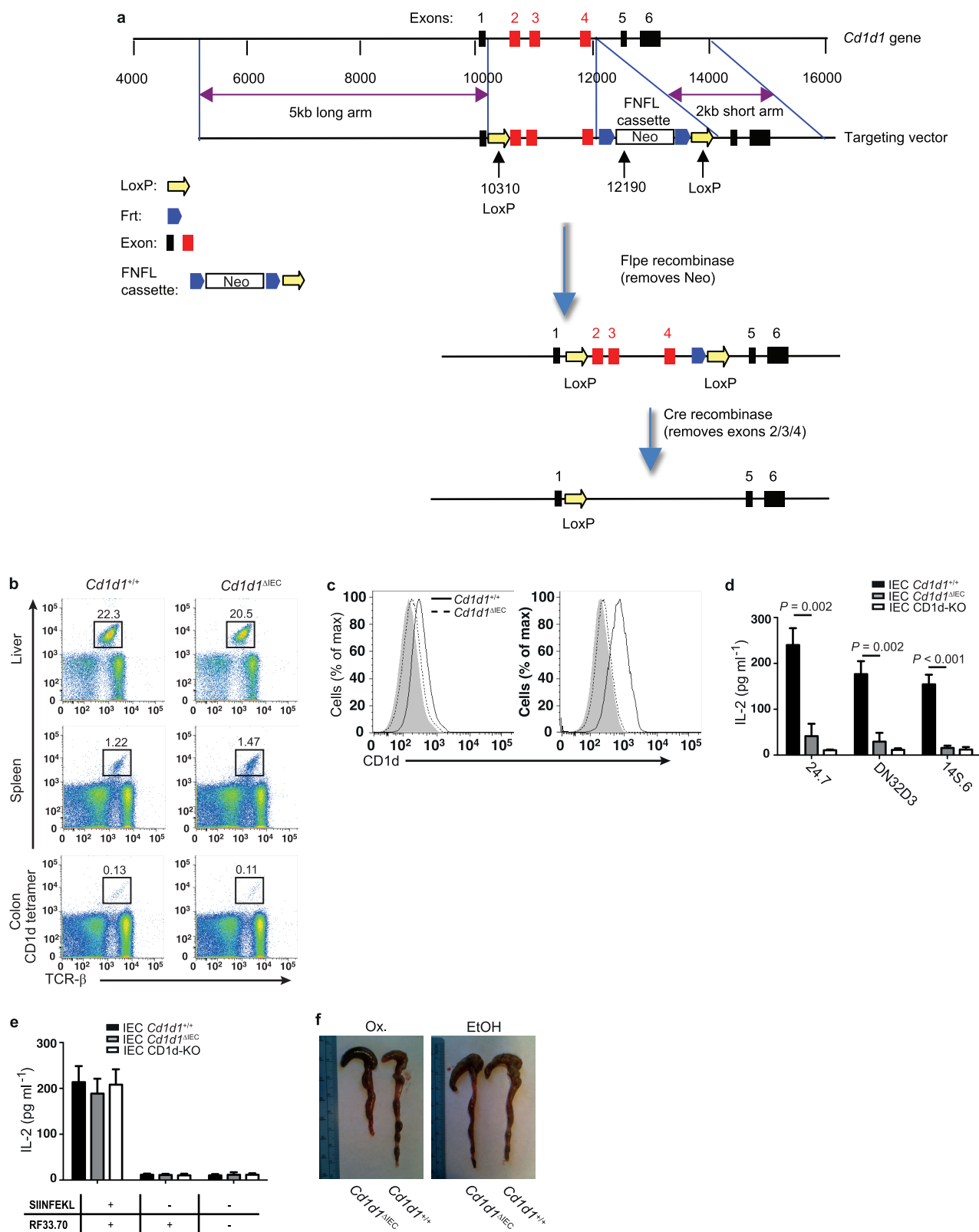


Extended Data Figure 3 | Impaired CD1d- but not MHC-class I-restricted antigen presentation in *Mttp*^{ΔIEC} mice. **a**, CD1d-mediated presentation of the exogenous lipid antigen α -GalCer to the invariant NKT-cell hybridomas 24.7 and DN32.D3 and of endogenous lipid antigens (autoreactivity) to the non-invariant NKT-cell hybridoma 14S.6 by IECs from CD1d-knockout (CD1d-KO) mice, *Mttp*^{ΔIEC} mice and wild-type littermates (*Mttp*^{+/+}) ($n = 5$ mice per group). **b**, Presentation of H2-K^b-restricted SIINFEKL by the indicated IECs (see **a**) to the SIINFEKL-responsive hybridoma RF33.70.

c, Representative histograms of CD1d cell surface expression as determined by flow cytometry of colonic IECs of the indicated mouse strains before (left, $n = 5$ mice per group) and 6 h after (right, $n = 5$ mice per group) rectal oxazolone challenge. **d**, Representative haematoxylin and eosin stainings of the indicated mouse strains upon rectal challenge with oxazolone (Ox.) or vehicle (ethanol (EtOH)). Scale bar, 40 μ m. Results representative of three independent experiments are shown. Mean \pm s.e.m. of triplicate cultures are shown. Student's *t*-test was applied.

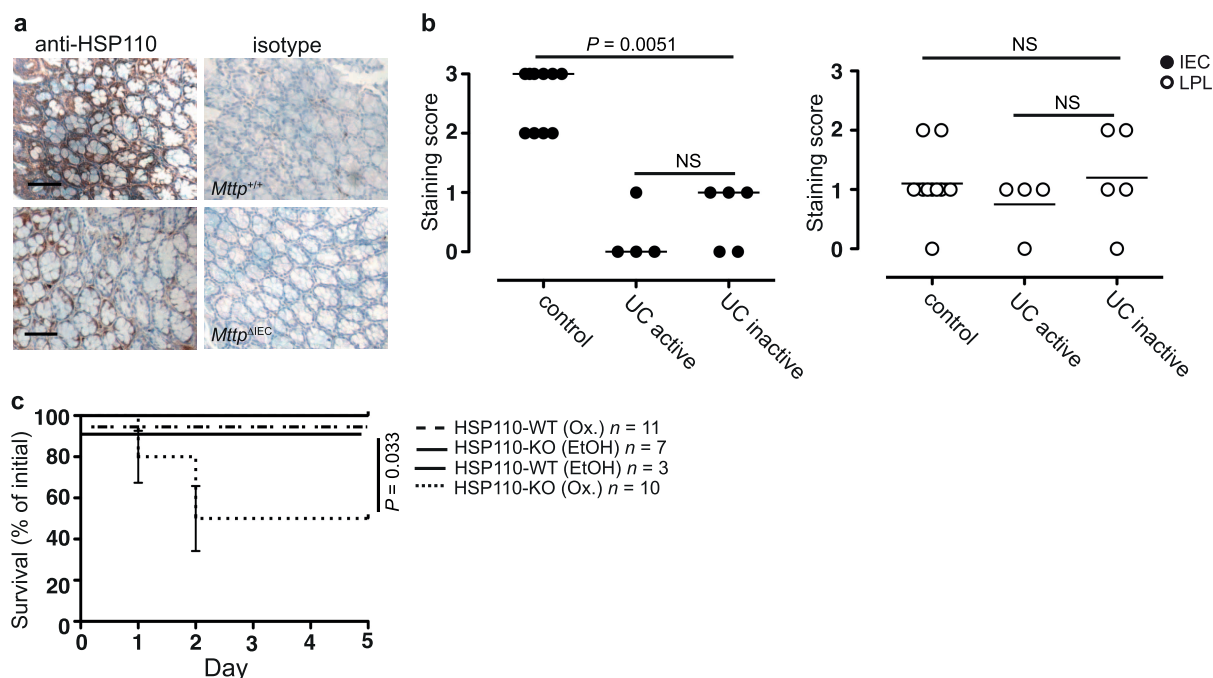


Extended Data Figure 4 | Increased morbidity and mortality in oxazolone-challenged $Mtp^{\Delta IEC}$ mice is due to CD1d-restricted components of the adaptive immune system. **a**, Intestinal permeability as determined by FD-4 before and 18 h after rectal challenge with oxazolone (ox.) in the indicated mouse strains. Each symbol represents a single mouse. **b**, IL-1 β secretion by CD11b $^{+}$ cells from colonic lamina propria of the indicated bone marrow chimaeras. CD11b $^{+}$ cells were isolated using magnetic microbeads 24 h after rectal challenge with oxazolone or ethanol and cultured for 24 h before measurement of IL-1 β in culture supernatants by ELISA. Mean \pm s.e.m. of triplicate cultures are shown. **c**, **d**, Survival and body weight of the indicated mouse strains at the indicated days after rectal oxazolone challenge. Results representative of three independent experiments are shown. Mean \pm s.e.m. of the indicated number of mice are shown. Student's t -test was applied.



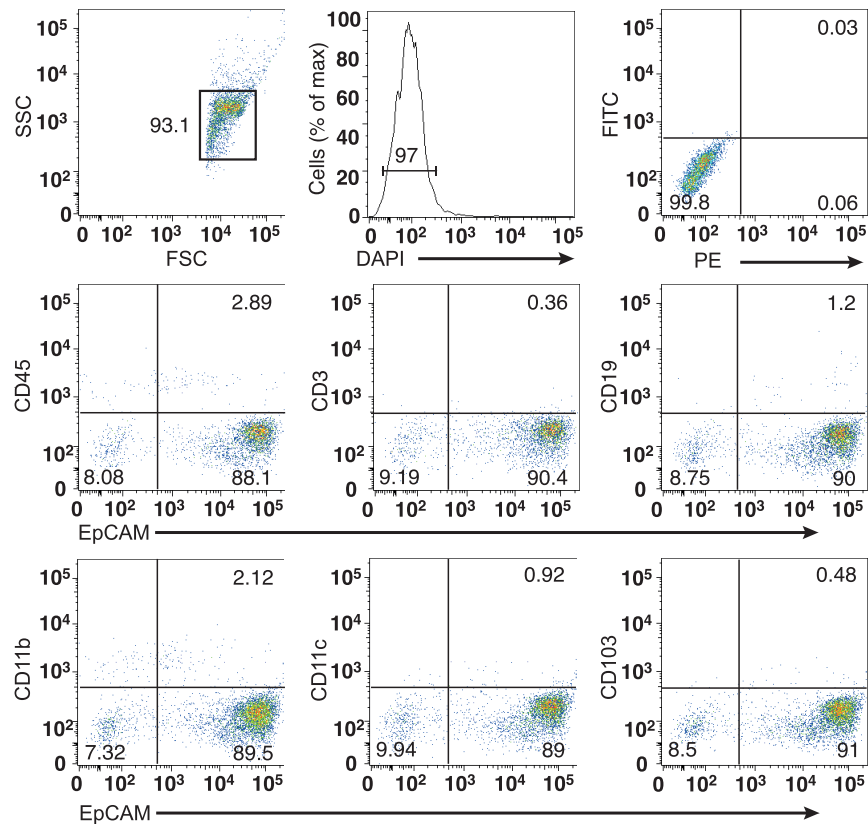
Extended Data Figure 5 | Development and characterization of mice with IEC-specific *Cd1d1* deletion. **a**, Schematic map of the targeting strategy for generation of *Cd1d1*^{fl/fl} mice. A LoxP (L83) site was inserted at 10310 and a FNFL (Frt-Neo-Frt-LoxP) cassette at 12190 to flank exons 2, 3, and 4 (about 1.9 kb) of the *Cd1d1* gene to generate the 'floxed/neo' *Cd1d1* allele. A gene-targeting vector was constructed by retrieving one 5 kb long homology arm (5' to L83), one 1.9 kb sequence containing exon 2/3/4, FNFL cassette, and one 2 kb short homology arm (end of FNFL to 3'). The FNFL cassette conferred G418 resistance during gene targeting in PTL1 (129B6 hybrid) embryonic stem (ES) cells. The targeted *Cd1d1* allele was PCR amplified and sequenced to confirm the targeted C57BL/6 allele based on the C57BL/6-specific mutation in the *Cd1d2* allele. Three targeted ES cells with targeted C57BL/6 allele were injected into C57BL/6 blastocysts to generate chimaeric mice. Male chimaeras were bred to bACTFpe females or EIIa-Cre females to transmit the floxed *Cd1d1* allele (*Cd1d1*^{fl/+}) (with neo cassette removed by FpI recombination) through the germ line. Mice carrying the floxed *Cd1d1* allele were crossed to tissue-specific VillinCre-ER^{T2} and exons 2, 3 and 4 and Frt-LoxP were removed by Cre recombinase in intestinal epithelial cells. **b–e**, Impaired CD1d- but not

MHC-class I-restricted antigen presentation in *Cd1d1*^{ΔIEC} mice. **b**, *Cd1d1*^{ΔIEC} mice exhibit normal numbers of invariant NKT cells in liver, spleen and colonic lamina propria as determined by flow cytometry using α -GalCer/CD1d tetramers ($n = 4$ mice per group). **c**, Representative histograms of CD1d cell surface expression on colonic IECs of the indicated mouse strains before (left, $n = 5$ mice per group) and 6 h after (right, $n = 5$ mice per group) rectal oxazolone challenge. **d**, CD1d-mediated antigen presentation by colonic IECs of the indicated mouse strains ($n = 5$ mice per group). Presentation of the model glycolipid antigen α -GalCer to invariant NKT-cell hybridomas 24.7 and DN32.D3 and of endogenous antigens (autoreactivity) to the non-invariant NKT-cell hybridoma 14S.6 is shown. **e**, Presentation of H2-K^b-restricted SIINFEKL presentation by the indicated IECs ($n = 5$ mice per group) to the SIINFEKL-responsive hybridoma RF33.70. Results representative of three independent experiments are shown **b–e**, Mean \pm s.e.m. of triplicate cultures are shown. Student's *t*-test was applied. **f**, Intestinal epithelial CD1d is protective in oxazolone colitis. Representative macroscopic colon images of *Cd1d1*^{ΔIEC} and wild-type littermates upon rectal challenge with oxazolone (Ox., $n = 5$ mice per group) or vehicle (ethanol (EtOH), $n = 4$ mice per group).



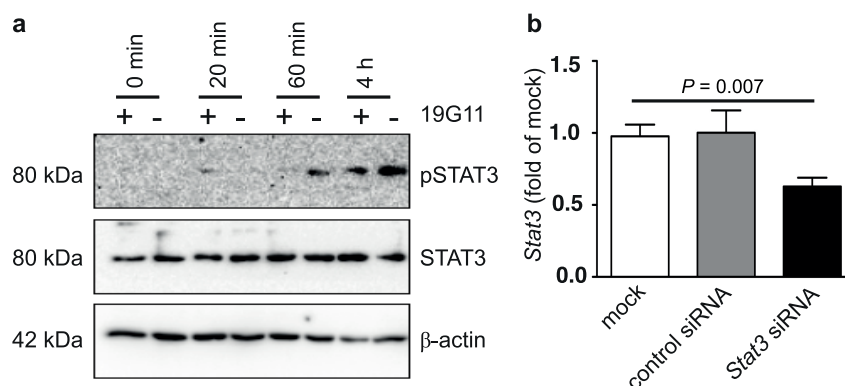
Extended Data Figure 6 | Epithelial HSP110 expression is decreased in *Mttp^{AIEC}* mice and in human ulcerative colitis. **a**, HSP110 immunohistochemistry in mice with IEC *Mttp* deletion (bottom) as compared with wild-type littermates (top) 6 h after rectal oxazolone challenge. **b**, HSP110 immunohistochemistry in the colonic intestinal epithelium and the lamina propria of healthy controls and patients with active and inactive ulcerative colitis. Signal intensity was scored on a scale from 0 to 3 in a blinded fashion.

Each symbol represents a single patient. The median and significance level as determined by the Mann–Whitney *U*-test are shown. **c**, Mortality in conventional HSP110-knockout mice in the oxazolone colitis model. Results representative of three independent experiments are shown. Mean \pm s.e.m. of the indicated number of mice is shown. Student's *t*-test and the log-rank test (survival) were applied. NS, not significant.



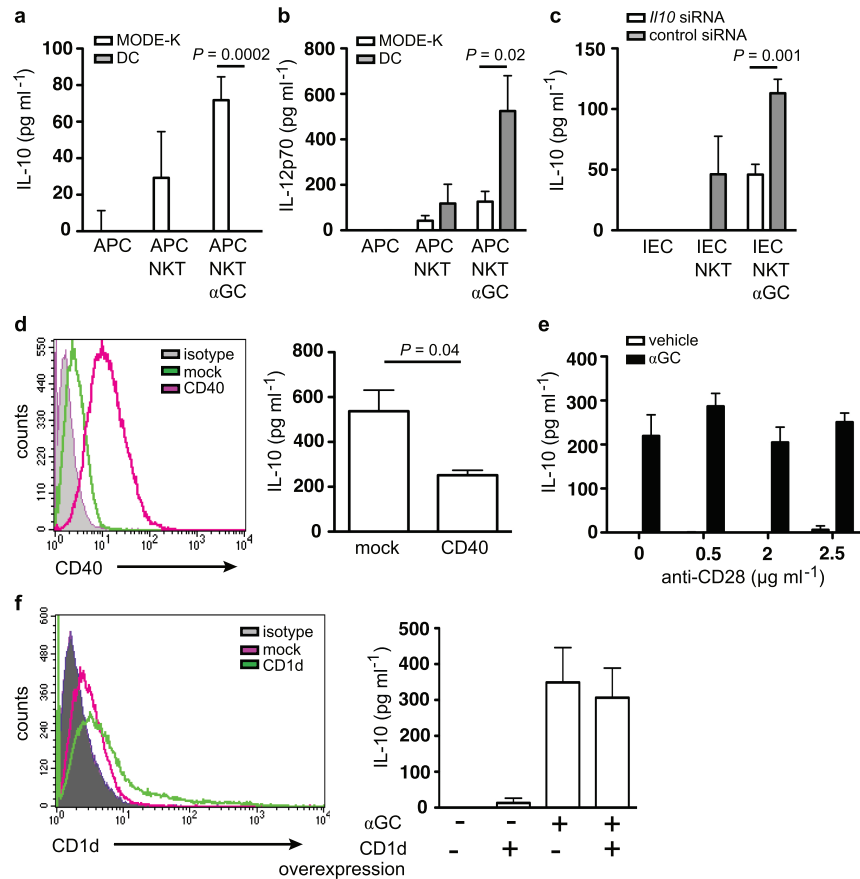
Extended Data Figure 7 | Purity of isolated IECs. After isolation of colonic IECs, potential contamination with haematopoietic cells was investigated by flow cytometry. The major population of cells was gated (upper left), was shown to contain viable cells (upper middle), and to contain largely EpCAM-positive

epithelial cells that do not stain with leukocyte markers (middle and bottom). The upper right panel shows isotype control staining. Results representative of three independent experiments ($n = 5$ mice per experiment) are shown.



Extended Data Figure 8 | CD1d-dependent STAT3 phosphorylation upon co-culture of IECs and iNKT cells. **a**, STAT3 phosphorylation in IECs upon co-culture with iNKT cells is CD1d-dependent. Expression of pSTAT3, STAT3 and β -actin as determined by western blotting of MODE-K cells at the indicated time after addition of the iNKT-cell hybridoma 24.7. Where indicated, MODE-K cells were pre-incubated with a monoclonal blocking

antibody directed against CD1d (19G11; +) or an isotype control (-). **b**, siRNA-mediated knockdown of *Stat3* in MODE-K cells as determined 68 h after siRNA transfection. Results representative of three independent experiments are shown. Means \pm s.e.m. of triplicates are shown. Student's *t*-test was applied.



Extended Data Figure 9 | CD1d-mediated cytokine production is dependent on the APC and expression of co-stimulatory molecules. a–f, IL-10 (a, c–f) and IL-12p70 (b) secretion of co-cultures of the IEC line MODE-K or splenic dendritic cells (DCs) together with the iNKT-cell hybridoma 24.7. APCs were loaded with α-GalCer (αGC; 100 ng ml⁻¹) before washing and co-culture with iNKT cells. c, MODE-K cells (IEC) were transfected with control siRNA or siRNA directed against *Il10* before co-culture with iNKT cells.

d, f, MODE-K cells were transfected with CD40 (d) and CD1d (f) as indicated. Histograms demonstrated increased expression of CD40 (d) and CD1d (f) after transfection. e, Stimulatory anti-CD28 antibody was added during co-culture of MODE-K and iNKT cells. Means ± s.e.m. of quadruplicate cultures are shown. Student's *t*-test was applied. Results are representative of two independent experiments.

Extended Data Table 1 | Genes downregulated in expression in oxazolone-challenged *Mttr*^{ΔIEC} mice

Fold Change*	GeneID	Description
8.721	<i>Ifi202b</i>	interferon activated gene 202B
3.878	<i>Eya3</i>	eyes absent 3 homolog (Drosophila)
3.7	<i>Sipa1l1</i>	signal-induced proliferation-associated 1 like 1
3.661	<i>Gna14</i>	AV230778 RIKEN full-length enriched, 0 day neonate skin Mus musculus cDNA clone 4632401H08 3', mRNA sequence.
3.534	<i>Cd79b</i>	CD79B antigen
3.347	<i>Mapk14</i>	mitogen activated protein kinase 14
3.347	<i>Aasdhpt</i>	aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase
3.259	<i>Marcks</i>	myristoylated alanine rich protein kinase C substrate
3.216	<i>Tgtp</i>	T-cell specific GTPase
3.216	<i>RbmX</i>	RNA binding motif protein, X chromosome
3.178	<i>Tcl1b3</i>	T-cell leukemia/lymphoma 1B, 3
3.162	<i>Clu</i>	clusterin
3.134	<i>Indo</i>	indoleamine-pyrrole 2,3 dioxygenase
3.12	<i>Ncl</i>	uy94h11.x1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3667269 3' similar to SW:NUCL_MOUSE P09405 NUCLEOLIN ; mRNA sequence.
3.101	<i>Hspa4</i>	heat shock protein 4
3.068	<i>Gbp2</i>	guanylate nucleotide binding protein 2
3.057	<i>Ncbp2</i>	nuclear cap binding protein subunit 2
3.034	<i>Mcm5</i>	minichromosome maintenance deficient 5, cell division cycle 46 (S. cerevisiae)
3.024	<i>Hsph1</i>	heat shock protein 110

* Fold change: *Mttr*^{+/+} / *Mttr*^{ΔIEC}Genes downregulated by more than threefold in intestinal mucosal scrapings of *Mttr*^{ΔIEC} mice compared with wild-type littermates 6 h after rectal challenge with oxazolone.

Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid

Long N. Nguyen¹, Dongliang Ma², Guanghou Shui³, Peiyan Wong², Amaury Cazenave-Gassiot³, Xiaodong Zhang², Markus R. Wenk³, Eyleen L. K. Goh² & David L. Silver¹

Docosahexaenoic acid (DHA) is an omega-3 fatty acid that is essential for normal brain growth and cognitive function^{1–4}. Consistent with its importance in the brain, DHA is highly enriched in brain phospholipids^{5–7}. Despite being an abundant fatty acid in brain phospholipids, DHA cannot be *de novo* synthesized in brain and must be imported across the blood–brain barrier, but mechanisms for DHA uptake in brain have remained enigmatic. Here we identify a member of the major facilitator superfamily—Mfsd2a (previously an orphan transporter)—as the major transporter for DHA uptake into brain. Mfsd2a is found to be expressed exclusively in endothelium of the blood–brain barrier of micro-vessels. Lipidomic analysis indicates that *Mfsd2a*-deficient (*Mfsd2a*-knockout) mice show markedly reduced levels of DHA in brain accompanied by neuronal cell loss in hippocampus and cerebellum, as well as cognitive deficits and severe anxiety, and microcephaly. Unexpectedly, cell-based studies indicate that Mfsd2a transports DHA in the form of lysophosphatidylcholine (LPC), but not unesterified fatty acid, in a sodium-dependent manner. Notably, Mfsd2a transports common plasma LPCs carrying long-chain fatty acids such as LPC oleate and LPC palmitate, but not LPCs with less than a 14-carbon acyl chain. Moreover, we determine that the phosphor-zwitterionic headgroup of LPC is critical for transport. Importantly, *Mfsd2a*-knockout mice have markedly reduced uptake of labelled LPC DHA, and other LPCs, from plasma into brain, demonstrating that Mfsd2a is required for brain uptake of DHA. Our findings reveal an unexpected essential physiological role of plasma-derived LPCs in brain growth and function.

Immunolocalization of Mfsd2a indicates that it is highly enriched in brain micro-vessels, where it is exclusively found in endothelium constituting the blood–brain barrier (BBB) (Fig. 1a, b, c, and Extended Data Fig. 1a, b), but not expressed in pericytes that enwrap the endothelium^{8,9} (Extended Data Fig. 1c, d). This localization pattern in the BBB was also noted in monkey (Extended Data Fig. 1e–g). Mfsd2a was found to be expressed in the BBB at embryonic day 15.5 (E15.5) (Extended Data Fig. 2a). This pattern of localization suggests a transport function in the BBB.

Male and female *Mfsd2a*-deficient (*Mfsd2a*-knockout) mice were born at Mendelian ratios, but had significantly increased postnatal mortality early in life¹⁰. Consistent with this, fetal and placental weights of knockout mice at E18.5 were similar to wild-type littermates (Extended Data Fig. 2b). In addition, knockout mice after weaning exhibited motor dysfunction¹⁰ with front-paw claspings during tail suspension (Extended Data Fig. 2c). Brain size and weight of knockout mice was also significantly lower for knockout mice than wild-type littermates (Extended Data Fig. 3a, b). However, there were no visible differences in gross anatomy of brains from knockout and wild-type mice (Extended Data Fig. 3c, d). Interestingly, cerebellum of knockout mice exhibited a significant loss of Purkinje cells (Fig. 1d, e). Furthermore, there was a significant decrease in neuronal cell density in the hippocampus, particularly in the CA1 and CA3 regions, but normal cell density in the dentate gyrus of knockout mice (Fig. 1f, g). Behavioural tests indicated that knockout mice had

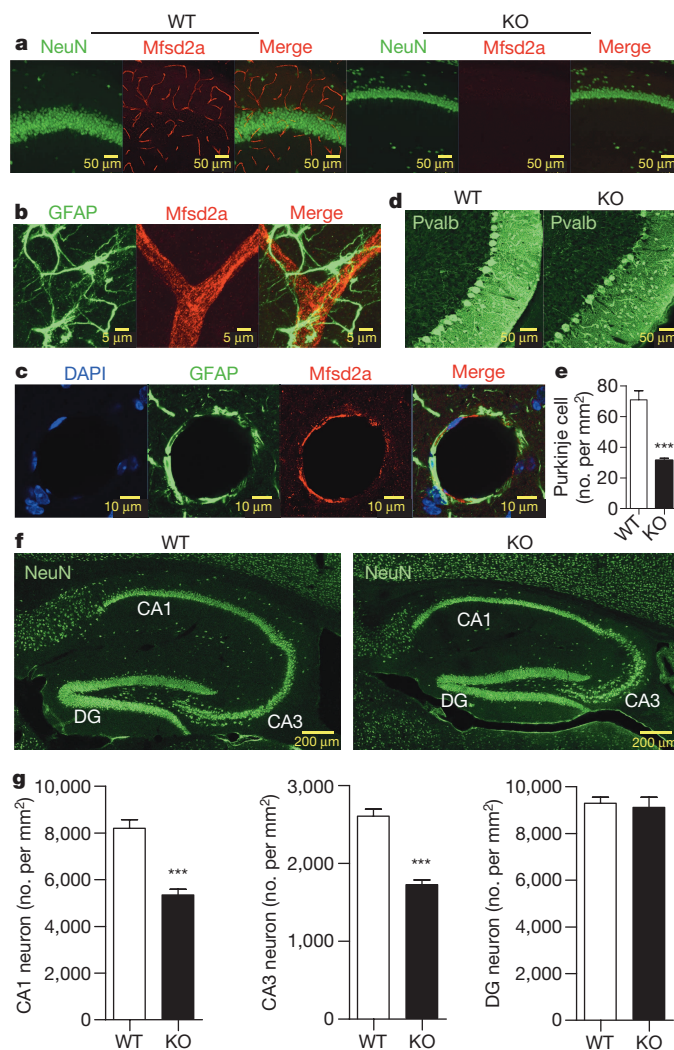


Figure 1 | Localization of Mfsd2a in the blood–brain barrier and neuronal deficits in *Mfsd2a*-knockout mice. **a**, Mfsd2a (red) is highly enriched in brain micro-vessels, shown here in CA1 region with NeuN staining (green) of mature neurons. **b**, **c**, Mfsd2a is expressed in endothelial cells of the blood–brain barrier in close contact with endfeet of astrocytes as shown by GFAP staining. **d**, Loss of Purkinje cells detected by parvalbumin (Pvalb) staining in cerebellum of knockout (KO) mice. **e**, Quantification of Purkinje cells in the cerebellum of wild-type (WT) and knockout mice. **f**, NeuN staining in hippocampus of sagittal brain sections of 8-week-old wild-type and knockout mice indicates decreased mature neurons in specific hippocampal regions of knockout mice. **g**, Quantification of neuron numbers in CA1, CA3 and dentate gyrus (DG) regions from mice examined in **f** above. *** $P < 0.001$. Data are expressed as mean \pm s.e.m.

¹Signature Research Program in Cardiovascular and Metabolic Disorders, Duke-NUS Graduate Medical School Singapore, 8 College Road, 169857 Singapore. ²Signature Research Program in Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School Singapore, 8 College Road, 169857 Singapore. ³Department of Biochemistry, National University of Singapore, 8 Medical Drive, Block MD7, 117597 Singapore.

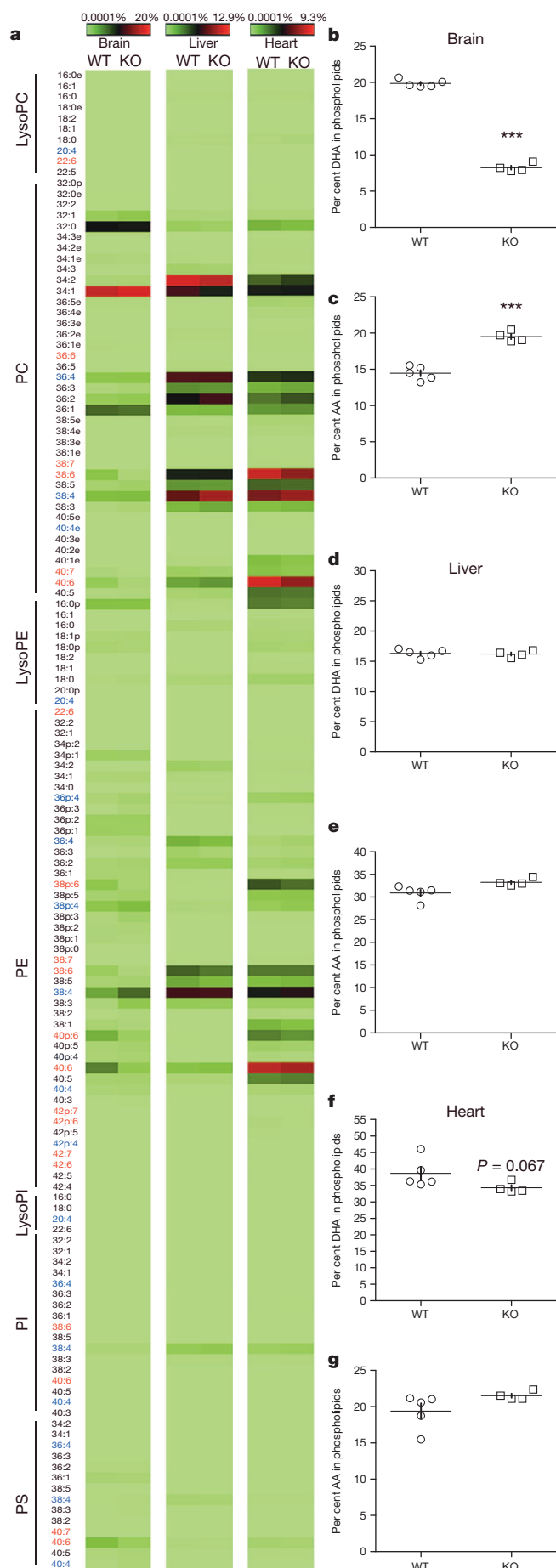


Figure 2 | Brains of *Mfsd2a*-knockout mice are DHA-deficient.

A comprehensive lipidomic analysis of brain, liver and heart phospholipids of adult *Mfsd2a*-knockout and wild-type mice. **a**, Heatmap representation of percentage of individual phospholipid species measured from brain, liver and heart. DHA- and arachidonic acid (AA)-containing species are highlighted in red and blue, respectively. e, ether; p, plasmalogen; LysoPC, lysophosphatidylcholine; LysoPI, lysophosphatidylinositol; LysoPE, lysophosphatidylethanolamine; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; PS, phosphatidylserine. **b–g**, Total DHA levels in brain, liver and heart phospholipids (**b, d, f**; wild-type, $n = 5$; knockout, $n = 4$), and total AA levels in brain, liver and heart (**c, e, g**). *** $P < 0.0001$ DHA and AA levels are expressed as mean \pm s.e.m. of the percentage of the total level of phospholipids (see the Source Data associated with this figure for the full data set).

severe deficits in learning, and short- and long-term memory, as well as severe anxiety (Extended Data Fig. 4).

The cognitive and behavioural dysfunction of *Mfsd2a*-knockout mice is reminiscent of omega-3 fatty-acid deficiency^{11,12}. We performed a comprehensive lipidomic analysis of brain, liver and heart by mass spectrometry on wild-type and knockout mice. Notably, we found that DHA, but not other fatty-acid species, in major phospholipid species of phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI) and phosphatidylserine (PS) of brain of knockout mice were significantly decreased compared to wild-type mice (Fig. 2a). DHA is found mainly as phospholipid species 38:6 and 40:6 in PE, PC, PI and PS¹³ (Fig. 2a). The total level of DHA-containing species was not significantly different in the liver and heart of knockout mice compared to wild-type mice (Fig. 2d, f). However, the total level of DHA in brain of knockout mice was reduced by 58.8% with minor changes in other fatty-acid species (Fig. 2b). Brains of knockout mice had a 33.8% increase in arachidonic acid in phospholipids (Fig. 2c), which is commonly increased in rodent models of DHA deficiency¹⁴. Arachidonic acid levels in liver and heart were unchanged (Fig. 2e, g). It is noteworthy that knockout mice were grown on DHA sufficient diet underscoring a physiological role of *Mfsd2a* in maintaining brain DHA levels. Despite the apparent lack of anatomical changes in the developing brains of *Mfsd2a*-knockout embryos, biochemical changes in brain phospholipids were apparent at E18.5 with significantly reduced levels of DHA in phospholipids (Extended Data Fig. 2d), indicating that *Mfsd2a* plays an essential role in maintaining DHA levels during embryogenesis.

The brain selectively accumulates a high amount of plasma-derived DHA. In plasma, the exchangeable pool of plasma DHA is found in albumin as either unesterified fatty acid or as LPC¹⁵. Using cell based assays, we found that *Mfsd2a* did not transport unesterified DHA, or other unesterified fatty acids (Extended Data Fig. 5a, b, and Supplementary Table 1). Remarkably, cells expressing *Mfsd2a* exhibited an enhanced concentration-dependent uptake of LPC [¹⁴C]DHA relative to control cells (Fig. 3a), indicating that *Mfsd2a* is indeed a LPC DHA transporter. Alanine mutagenesis of the phylogenetically conserved residues aspartate 92 (Asp92Ala) and 96 (Asp96Ala) critical for sodium binding in MFS proteins¹⁶ resulted in reduced and absence of transport, respectively (Fig. 3a). Asp92Ala and Asp96Ala mutants had similar expression as wild type (Extended Data Fig. 5c, d). We next examined transport specificity of *Mfsd2a* for the most common LPCs found in plasma, LPC oleate and LPC palmitate. Cells expressing wild-type *Mfsd2a* showed concentration-dependent uptake of LPC [¹⁴C]oleate and LPC [³H]palmitate (Fig. 3b, c). Human *Mfsd2a* also transported LPCs in a similar manner (Extended Data Fig. 5e–h). *Mfsd2a* had highest capacity to transport LPC DHA, followed by LPC oleate, and LPC palmitate (Fig. 3d). Moreover, cells expressing wild-type and partially active Asp92Ala mutant exhibited saturation kinetics over time (Extended Data Fig. 5i). Importantly, *Mfsd2a*-dependent transport of LPC [¹⁴C]DHA and LPC [¹⁴C]oleate resulted in their rapid conversion into phosphatidylcholines (Fig. 3e–h). Moreover, transport by *Mfsd2a* resulted in increases in phosphatidylcholine mass, indicating net uptake of LPC into cells (Extended Data Fig. 5j).

LPC transport by *Mfsd2a* was sodium-dependent (Fig. 3i). Consistent with the characteristics of other sodium-dependent MFS symporters,

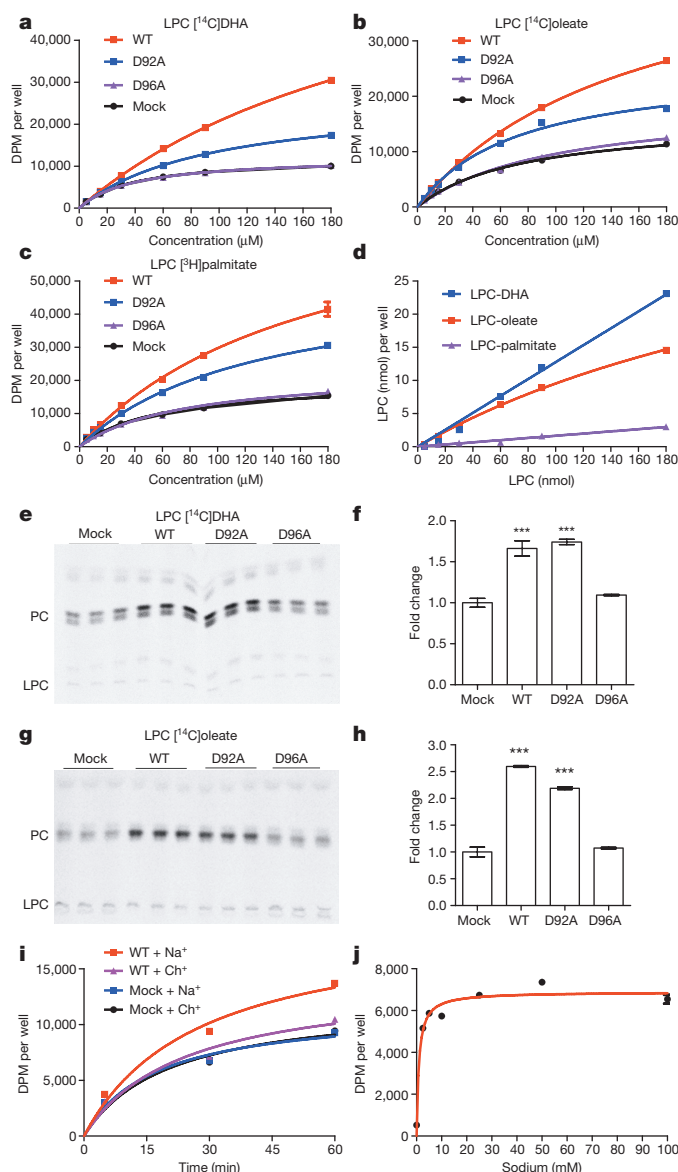


Figure 3 | Cell-based transport assays of radiolabelled LPCs.

a–c, Concentration-dependent transport of LPC [^{14}C]DHA, LPC [^{14}C]oleate and LPC [^3H]palmitate after 30 mins. Mouse Mfsd2a (wild-type) and mutant constructs D92A and D96A were tested for uptake of radiolabelled LPCs at indicated concentrations. **d**, Comparison of transport preference of LPC [^{14}C]DHA, LPC [^{14}C]oleate and LPC [^3H]palmitate. **e**, Biological incorporation of radiolabelled LPC [^{14}C]DHA. **f**, Quantification of PC bands from thin-layer chromatography (TLC) plates shown in **e**. **g**, LPC [^{14}C]oleate bio-incorporation into phosphatidylcholine. **h**, Quantification of PC bands from TLC plates shown in **g**. **i**, Transport activity of mouse Mfsd2a is dependent on sodium. Ch^+ , choline. **j**, Dose-response curve for sodium-concentration dependency of transport of 50 μM LPC [^{14}C]oleate by Mfsd2a. *** $P < 0.0001$. For **a–d**, **f**, **h–j**, data in triplicates are expressed as mean \pm s.e.m.

the transport is highly sensitive to low sodium concentrations, indicating high affinity for sodium (Fig. 3j)¹⁷. LPC transport was not dependent on pH or lithium (Extended Data Fig. 5k, l). In addition, both LPC [^3H] 16:0 solubilized in ethanol or in micelles was transported by Mfsd2a, albeit with lesser capacity compared with bovine serum albumin (BSA)-bound form, indicating that albumin is not essential for transport (Extended Data Fig. 5m). Our findings reveal Mfsd2a as the first facilitative transporter identified for the transport of phospholipids.

We next sought to determine ligand specificity and the chemical features of the LPC ligands that are required for transport by Mfsd2a. To carry out this goal, we set up competition assays using cells expressing

Mfsd2a or mock control cells, and treated them with 25 μM LPC [^3H] palmitate in the presence or absence of a tenfold molar excess of unlabelled competitor. We found that a LPC with a minimum acyl chain length of 14 carbons can effectively compete for uptake (Extended Data Fig. 6a–c). The lysophospholipids LPE and LPS showed weak competition, whereas LPA was non-competitive for the transport of LPC [^3H] palmitate (Extended Data Fig. 6c). We confirmed that Mfsd2a can directly transport LPE using fluorescent LPE (Extended Data Fig. 7).

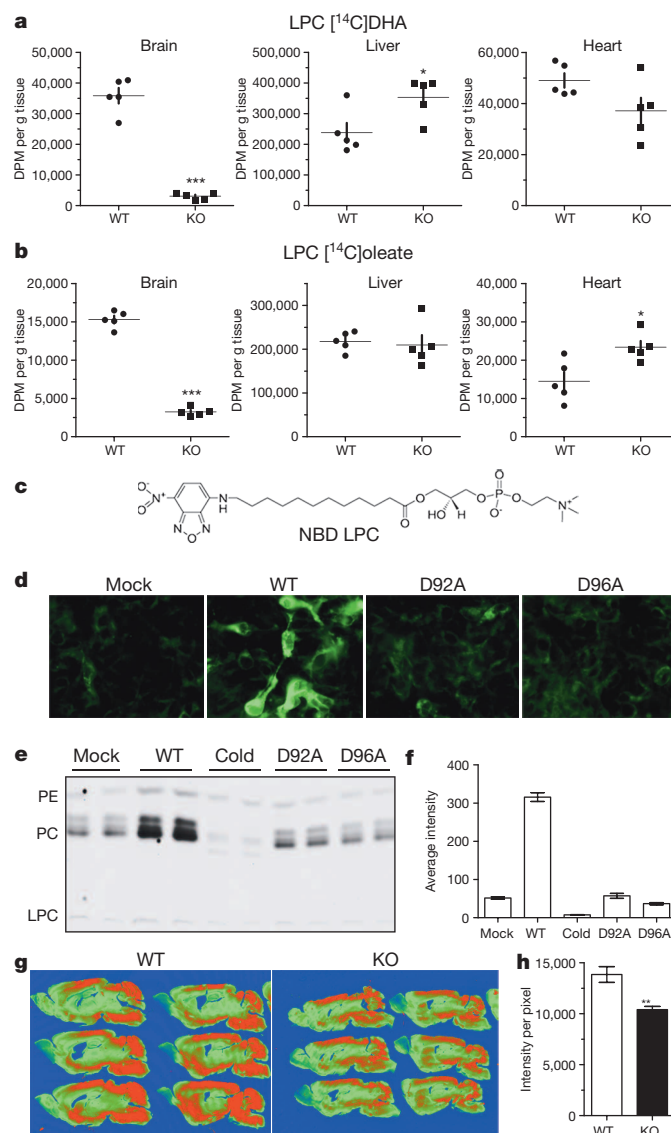


Figure 4 | Uptake of radiolabelled LPCs by brain is decreased in

Mfsd2a-knockout mice. **a, b**, Male mice of 6–7 weeks of age were injected intravenously with the same dose of LPC [^{14}C]DHA (**a**) and LPC [^{14}C]oleate (**b**). Brain, liver and heart were collected 2 h post injection for lipid extraction and DPM was quantified using scintillation counting. Uptake is expressed as mean \pm s.e.m. (wild-type, $n = 5$; knockout, $n = 5$). **c**, Structure of fluorescence NBD LPC. **d**, HEK293 cells expressing wild-type Mfsd2a showed significantly enhanced uptake activity to NBD LPC compared with mock (empty plasmid), D92A and D96A mutant. **e**, TLC analysis showed that NBD LPC was bio-incorporated into PC. Transport of NBD LPC was suppressed by tenfold molar excess of LPC 18:0 (cold). Mutant D92A, D96A had similar transport activity for NBD LPC as mock-transfected cells. **f**, Quantification of PC band from TLC plates shown in **e**. **g**, Brain uptake of NBD LPC was decreased in knockout mice. Male mice (wild-type, $n = 3$; knockout, $n = 3$) aged 6 weeks old were intravenously injected with 300 μg NBD LPC-BSA complex. **h**, Fluorescence from 15 brain sections of wild-type and knockout mice was quantified and expressed as fluorescence intensity per pixel. *** $P < 0.0001$, ** $P < 0.001$, * $P < 0.05$. Data are expressed as mean \pm s.e.m.

These results indicate that the zwitterionic charge of the phosphatidylcholine headgroup is critical for ligand transport. Furthermore, glycerophosphatidylcholine was not a competitor (Extended Data Fig. 6c), supporting the conclusion that a long acyl chain of the LPC is a requisite for ligand transport. Lysoplasmalogen, lysoplatelet activating factor, and platelet activating factor were strong competitors, indicating that the carbonyl group of the acyl chain of LPC is not required for transport (Extended Data Fig. 6d–f). Lysosphingomyelin also competed for LPC [³H]palmitate transport, indicating that the glycerol backbone is not required for ligand function (Extended Data Fig. 6f), further supporting the conclusion that the choline portion of the phosphocholine headgroup is essential for ligand transport. Using lysolipid-like detergents we confirmed that an alkyl side chain of a minimal length of 14 carbons and the phosphocholine headgroup, but not the glycerol backbone and carbonyl group, are essential chemical features of the LPC ligand (Extended Data Fig. 6g, h).

Remarkably, *Mfsd2a*-knockout mice had a reduction in brain uptake of LPC [¹⁴C]DHA and LPC [¹⁴C]oleate by 90% and 80%, respectively (Fig. 4a). LPC [¹⁴C]DHA or LPC [¹⁴C]oleate uptake in peripheral tissues of knockout mice were not decreased (Fig. 4a, b). Brain uptake of unesterified [¹⁴C]DHA was significantly lower relative to LPC [¹⁴C]DHA uptake, but was not reduced in knockout relative to wild-type mice (Extended Data Fig. 8), indicating brain accretion of DHA is mainly through LPC DHA uptake via *Mfsd2a*.

In a separate approach to examining *in vivo* transport of LPCs by *Mfsd2a*, we tested the transport of nitrobenoxadiazole (NBD) LPC, a fluorescent LPC analogue (Fig. 4c). Similar to native LPCs, NBD LPC was transported by *Mfsd2a* and bio-incorporated into phospholipids (Fig. 4d–f). Intravenous injection of NBD LPC into knockout and wild-type mice resulted in a significant reduction in the fluorescence in knockout brain sections (Fig. 4g, h). Similar findings were obtained using TopFluor LPC, a fluorophore that is structurally different to NBD LPC (Extended Data Fig. 9).

Dietary DHA treatment did not rescue knockout phenotypes, such as brain size and anxiety (Extended Data Fig. 10a–e). Lack of rescue can be explained by a dramatic reduction in embryonic brain uptake of dietary DHA at E18.5 (Extended Data Fig. 10f), consistent with reduced embryonic brain DHA levels (Extended Data Fig. 2d).

Taken together, the current study identifies the orphan transporter *Mfsd2a* as a sodium-dependent LPC symporter, and represents the major mechanism by which DHA enters the brain, and indicates for the first time an important physiological role of plasma-derived LPCs for normal brain growth and function. Given these findings, we propose to rename *Mfsd2a* to sodium-dependent LPC symporter 1 (NLS1).

METHODS SUMMARY

Animals. *Mfsd2a*-knockout mice were generated as described previously¹⁰. Experimental protocols were approved by the SingHealth Institutional Animal Care and Use Committee.

Lipidomic analysis. Phospholipid species were measured by high performance liquid chromatography (HPLC) 1100 system (Agilent) coupled with a Triple Quadrupole Ion Trap mass spectrometer (4000Qtrap, Applied Biosystem). Levels of individual phospholipid species were analysed as described previously¹⁸.

***In vitro* transport of radiolabelled and fluorescent LPCs.** Uptake assays of radiolabelled LPCs or fluorescent LPCs were tested using HEK293 cells overexpressing *Mfsd2a* and mutants constructs. For sodium dependent assay, radiolabelled LPCs were diluted in transport buffer (5 mM KCl, 10 mM Hepes, pH 7.4) with 150 mM NaCl or 150 mM choline chloride.

***In vivo* transport of radiolabelled, fluorescent LPCs and unesterified DHA.** Male and female mice were intravenously injected with 1 μmol of radiolabelled LPC–BSA complex. At 2 h post injection, mice were perfused with PBS containing 0.5% BSA. Tissues were collected for lipid extraction, and scintillation counted. Similar experiments were also performed with an injection of 300 μg NBD LPC–BSA complex or 300 μg TopFluor LPC–BSA complex per mouse. To study fetal transport, pregnant females at E18.5 was gavaged with bolus of [¹⁴C]DHA–BSA complex. Brains of fetuses were collected 20 h post gavage and weighed, and lipid extracts were scintillation counted.

Statistical analysis. Statistical differences of DHA and amino acid levels, and histological analyses between genotypes were calculated using an unpaired Student's

t-test. Statistical analysis of LPC [¹⁴C]DHA, LPC [³H]palmitate, LPC [¹⁴C]oleate signals were calculated using two-way analysis of variance (ANOVA); a *P* < 0.05 was considered to be significant. For the behavioural tests, a one-way ANOVA was used to analyse the zero maze, light–dark box, open field and Y maze. A two-way ANOVA was used to analyse novel object test. Bonferroni corrected pair-wise comparisons were used as post-hoc tests. Data were expressed as mean ± s.e.m. and *P* < 0.05 was considered statistically significant.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2013; accepted 10 March 2014.

Published online 14 May 2014.

- Kidd, P. M. Omega-3 DHA and EPA for cognition, behavior, and mood: clinical findings and structural-functional synergies with cell membrane phospholipids. *Altern. Med. Rev.* **12**, 207–227 (2007).
- Horrocks, L. A. & Yeo, Y. K. Health benefits of docosahexaenoic acid (DHA). *Pharmacol. Res.* **40**, 211–225 (1999).
- Mozaffarian, D. & Wu, J. H. Omega-3 fatty acids and cardiovascular disease: effects on risk factors, molecular pathways, and clinical events. *J. Am. Coll. Cardiol.* **58**, 2047–2067 (2011).
- Connor, W. E. Importance of n-3 fatty acids in health and disease. *Am. J. Clin. Nutr.* **71**, 171S–175S (2000).
- Breckenridge, W. C., Gombos, G. & Morgan, I. G. The lipid composition of adult rat brain synaptosomal plasma membranes. *Biochim. Biophys. Acta* **266**, 695–707 (1972).
- Innis, S. M. Dietary (n-3) fatty acids and brain development. *J. Nutr.* **137**, 855–859 (2007).
- Salem, N., Jr, Litman, B., Kim, H. Y. & Gawrisch, K. Mechanisms of action of docosahexaenoic acid in the nervous system. *Lipids* **36**, 945–959 (2001).
- Armulik, A. et al. Pericytes regulate the blood–brain barrier. *Nature* **468**, 557–561 (2010).
- Bell, R. D. et al. Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 409–427 (2010).
- Berger, J. H., Charron, M. J. & Silver, D. L. Major facilitator superfamily domain-containing protein 2a (MFS2A) has roles in body growth, motor function, and lipid metabolism. *PLoS ONE* **7**, e50629 (2012).
- Lafourcade, M. et al. Nutritional omega-3 deficiency abolishes endocannabinoid-mediated neuronal functions. *Nature Neurosci.* **14**, 345–350 (2011).
- Carrié, I., Clement, M., de Javel, D., Frances, H. & Bourre, J. M. Phospholipid supplementation reverses behavioral and biochemical alterations induced by n-3 polyunsaturated fatty acid deficiency in mice. *J. Lipid Res.* **41**, 473–480 (2000).
- Kim, H. Y. Novel metabolism of docosahexaenoic acid in neural cells. *J. Biol. Chem.* **282**, 18661–18665 (2007).
- Simopoulos, A. P. The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Exp. Biol. Med. (Maywood)* **233**, 674–688 (2008).
- Croset, M., Brossard, N., Polette, A. & Lagarde, M. Characterization of plasma unsaturated lysophosphatidylcholines in human and rat. *Biochem. J.* **345**, 61–67 (2000).
- Granell, M., Leon, X., Leblanc, G., Padros, E. & Lorenz-Fonfria, V. A. Structural insights into the activation mechanism of melibiose permease by sodium binding. *Proc. Natl Acad. Sci. USA* **107**, 22078–22083 (2010).
- Paroder-Belenitsky, M. et al. Mechanism of anion selectivity and stoichiometry of the Na⁺/I⁻ symporter (NIS). *Proc. Natl Acad. Sci. USA* **108**, 17933–17938 (2011).
- Shui, G. et al. Comparative plasma lipidome between human and cynomolgus monkey: are plasma polar lipids good biomarkers for diabetic monkeys? *PLoS ONE* **6**, e19731 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by grants from the Singapore Ministry of Health's National Medical Research Council CBRG/0012/2012 (to D.L.S.), the Singapore National Research Foundation Competitive Research Program grants 2008-01 (to E.L.G.) and 2007-04 (to M.R.W.), National University of Singapore's Life Sciences Institute (to M.R.W.), and Singapore National Medical Research Council Translational and Clinical Research Program NMRC/TCR/003-GMS/2008 (to X.Z.). We would like to thank B. Tan (Duke-NUS) for technical assistance with lipid extractions, and S. Ying (Duke-NUS) for assistance with behavioural phenotyping.

Author Contributions L.N.N. designed experiments; performed all *in vitro* experiments, *in vitro* and *in vivo* transport experiments in cells and mice, lipid extractions for lipidomic analysis, and fluorescence microscopy; analysed all data; and wrote the paper. D.M. performed immunolocalization studies and provided some technical support with mouse perfusions. G.S. performed lipidomic analysis. P.W. performed behaviour and learning and memory studies in mice. A.C.-G. performed lipidomic analysis. X.Z. supervised the behavioural core. M.R.W. supervised the lipidomic analysis. E.L.K.G. provided expertise with designing and interpreting immunolocalization studies. D.L.S. conceived and designed the study and experiments, performed *in vivo* transport experiments, analysed data, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L.S. (david.silver@duke-nus.edu.sg).

METHODS

Chemicals. Non-radiolabelled lysophosphatidylcholines, lysophosphatidylethanolamine, lysophosphatidylserine and other lysophospholipids were purchased from Avanti Polar Lipids. Radiolabelled 1-palmitoyl 2-lysophosphocholine (LPC [^3H]palmitate), 1-oleoyl 2-lysophosphocholine (LPC [^{14}C]oleate), 1-docosahexaenoyl 2-lysophosphocholine (LPC [^{14}C]DHA), and [^{14}C]docosahexaenoic acid ([^{14}C]DHA) were purchased from American Radiochemicals. Lysophospholipids either in chloroform (non-labelled) or ethanol:toluene (radiolabelled) were completely dried under the nitrogen gas and were solubilized in 12% fatty acid-free BSA (Sigma), which was dissolved in 150 mM NaCl. To prepare LPC [^{14}C]oleate and LPC oleate mixture, 25 μCi LPC [^{14}C]oleate (specific activity 55 mCi mmol^{-1}) were dried and dissolved in 3 ml 20 mM non-labelled LPC stearate-BSA. LPC [^3H]palmitate and LPC palmitate mixture were prepared by dissolved 25 μCi LPC [^3H]palmitate (specific activity 60 Ci mmol^{-1}) in 4 ml 10 mM LPC palmitate. 1-docosahexaenoyl LPC was prepared from hydrolysis of 1,2-didocosaheptaenoyl PC with honey bee venom PLA2 (Sigma) in borax buffer (pH 8.5) containing 5 mM CaCl_2 and purified by TLC methods. To prepare LPC [^{14}C]DHA and LPC DHA mixture, 10 μCi LPC [^{14}C]DHA was mixed with non-labelled LPC DHA-BSA complex to a final concentration of 10 mM. To prepare [^{14}C]DHA and DHA mixture, 50 μCi [^{14}C]DHA was mixed with non-labelled DHA-BSA complex to a final concentration of 12.2 mM. DHA oil (HiDHA) was a gift from NuMega.

Animals. *Mfsd2a*-knockout mice were generated as described previously¹⁰. Mice were maintained on a high-energy diet 5LJ5 (PicoLab) containing a total 11% fat. Pups were weaned at 3 weeks old and maintained on the high-energy diet. Experimental protocols were approved by the SingHealth Institutional Animal Care and Use Committee. Mice in all experiments were age- and gender-matched, and not randomized into groups. Experiment involving mice were not carried out with the experimenter blinded. Samples sizes used in animal experiments were based on empirical determination in preliminary experiments.

Lipidomic analysis. For tissue lipid analysis, brain, liver and heart of female wild-type and knockout mice aged 7–8 weeks born to *Mfsd2a* heterozygous (HET) mothers, and brains from E18.5 embryos were collected and immediately frozen with liquid nitrogen until extraction. Lipid extraction was followed by the chloroform-methanol method. For cell-based free fatty acid transport, lipids from HEK293 cells expressing mouse *Mfsd2a* and mock controls after overnight treatment with 100 μM docosahexaenoic, arachidonic, eicosapentaenoic, alpha-linolenic, linoleic, oleic, palmitic acids in BSA complex were extracted using HIP (hexane:iso-propanol 3:2, per volume) buffer and dried under N_2 gas. Phospholipid species were measured by high performance liquid chromatography (HPLC) 1100 system (Agilent) coupled with an Triple Quadrupole Ion Trap mass spectrometer (4000Qtrap, Applied Biosystem). Levels of individual phospholipid species were analysed using MRM as previously described¹⁸ and quantified using spiked internal standards that includes dimyristoyl PC, dimyristoyl PE, dimyristoyl PS (Avanti Polar Lipids, Alabaster, Alabama, USA), and dioctanoyl PI (Echelon Biosciences, Salt Lake City, Utah, USA).

In vitro transport of radiolabelled and fluorescent LPCs. Radiolabelled LPC palmitate (LPC [^3H]palmitate), oleate (LPC [^{14}C]oleate), and docosahexaenoate (LPC [^{14}C]DHA) or TopFluor LPC, TopFluor LPE, and NBD LPCs were dissolved in 12% BSA, which was diluted in 150 mM NaCl. Uptake assays of radiolabelled LPCs or fluorescent LPCs were tested using HEK293 cells overexpressing *Mfsd2a* and mutants constructs. In brief, HEK293 cells at 90–95% confluency were transfected using lipofectamine 2000 (invitrogen) with pcDNA3.1Mfsd2a (wild-type), pcDNA3.1Mfsd2aD92A (Asp92Ala), pcDNA3.1Mfsd2aD96A (Asp96Ala), or pcDNA3.1 (mock) plasmids. Uptake assays were performed after 24 h of transfections. Prior to ligand incubation, HEK293 transfected cells were washed with serum-free DMEM before assays. For concentration and time dependent assays, radiolabelled LPCs were diluted in pre-warmed DMEM media. For sodium dependent assay, radiolabelled LPCs were diluted in transport buffer (5 mM KCl, 10 mM Hepes, pH 7.4) with 150 mM NaCl or 150 mM choline chloride. For sodium concentration dependent assays, any reductions in the concentration of NaCl were replaced by choline chloride in order to maintain a constant cation molarity of 150 mM. All assays were performed in 12-well plates with triplicates and at 37 °C.

For albumin-free ligand transport assays. To prepare LPC palmitate dissolved in ethanol, 0.75 μCi LPC [^3H]palmitate were diluted in LPC palmitate in chloroform. Mixture was dried and dissolved in 50 μl ethanol before adding 6 ml of transport buffer with 150 mM sodium as described above to have 50 μM LPC palmitate. To prepare LPC palmitate micelles, 0.75 μCi LPC [^3H]palmitate were diluted in LPC palmitate in chloroform. Mixture was dried and dissolved in 6 ml of transport buffer with 150 mM sodium to have 100 μM LPC palmitate and sonicated on ice for 5 min. Activated charcoal was added and spun to remove the monomers of LPC palmitate. Transport assays were similar performed with HEK293 cells overexpressing with pcDNA3.1Mfsd2a or pcDNA3.1 plasmid as control for 30 mins at 37 °C.

Competition transport assay. In brief, 24 h post transfection of HEK293 cells with pcDNA3.1Mfsd2a (wild-type), pcDNA3.1Mfsd2aD92A (Asp92Ala), pcDNA3.1Mfsd2aD96A, or pcDNA3.1 (mock) plasmids were washed once with serum-free DMEM medium before addition of a mixture of 25 μM radiolabelled LPC palmitate and 250 μM cold competitors, which were dissolved in 12% BSA. Total BSA concentration was kept constant in samples with or without cold competitors. Assays were performed at 37 °C for 30 min. Competition assays with other LPC analogs such as foscholine detergents, and PAF were performed under the same conditions except for 15 min. The reduced reaction time was necessary in order to limit potential negative effects of detergents and bioactive lipids on cell survival. All assays were performed in 12-well plates with triplicates and at 37 °C.

In vivo transport of radiolabelled, fluorescent LPCs and unesterified DHA. Male and female mice of age of 6–8 weeks were intravenously injected with 75 μl of 20 mM radiolabelled LPC [^{14}H]oleate, 100 μl of 10 mM LPC [^{14}C]DHA-BSA complex or 82 μl of 12.2 mM [^{14}C]DHA-BSA complex in a total volume of 150 μl of PBS. At 2 h post injection, mice were anaesthetized and perfused for 5 min with PBS containing 0.5% BSA in order to remove blood and lipid tracer bound to the brain vasculature. Tissues were collected for lipid extraction. For lipid extraction, tissues were weighed and similar amounts of tissues were homogenized in chloroform:methanol. Lipids from the organic phase were mixed with scintillant and scintillation counted. Similar experiments were also performed with an injection of 300 μg NBD LPC-BSA complex or 300 μg TopFluor LPC-BSA complex per mouse. In this experiment, mice were perfused for 5 min with PBS followed 15 min perfusion with 4% paraformaldehyde (PFA) for tissue fixation. Brain sections of wild-type and knockout with a thickness of 40 μm were prepared and scanned using green fluorescence mode in Typhoon FLA 9000 scanner (Agilent). NBD LPC accumulation was expressed as fluorescence of each section per its area. Dietary DHA transport from mother to fetus was also measured using [^{14}C]DHA-BSA complex. Pregnant females at E18.5 was gavaged with bolus of [^{14}C]DHA-BSA complex (200 μl of 10 mM [^{14}C]DHA-BSA per mouse). Brains of fetuses were collected 20 h post gavage and weighed. Lipid extraction and radioactivity were performed as described above.

TLC analysis of phospholipids. HEK293 cells overexpressing with pcDNA3.1Mfsd2a (wild-type), pcDNA3.1Mfsd2aD92A (Asp92Ala), pcDNA3.1Mfsd2aD96A, or pcDNA3.1 (mock) plasmids or human *Mfsd2a* (in Sport6 plasmid) were washed once with serum-free DMEM medium before incubation with 50 μM radiolabelled LPC [^{14}C]oleate, LPC [^{14}C]DHA or with 50 μM TopFluor LPC and NBD LPC. Lipids were extracted twice with HIP buffer for 30 min. Lipids were dried with nitrogen stream and reconstituted in chloroform and spotted on TLC plates (Milipore). Solvent for phospholipid separation was a solution of chloroform, methanol and ammonia (25%) (50:25:6 per volume). Thin-layer chromatography (TLC) plates of radiolabelled phospholipids were dried for 30 min and exposed overnight to PhosphorScreens and scanned with Typhoon FLA 9000 scanner (Agilent). TLC plates of fluorescent phospholipids were scanned with Typhoon FLA 9000 scanner and quantified using Imagequant software.

Histological studies. Adult male wild-type and knockout mice aged 7.5–8 weeks old born to HET parents, were deeply anaesthetized and perfused transcardially with 50 ml of saline, followed by 100 ml of 4% PFA in 0.1 M PBS (pH 7.4) for 30 min. For embryos, brains were fixed in 4% PFA then 30% sucrose both overnight. Sagittal sections of 40 μm thickness were sectioned in a cryostat, and serial sections were transferred to different wells of a 24-well tissue-culture dish for immunostaining. Sections of embryos were embedded on slides. The brain sections were processed for immunocytochemistry procedure using antibodies against NeuN (1:1000, Chemicon, California, USA), *Mfsd2a* (1:500), GFAP (1:1000, Chemicon, California, USA), parvalbumin (1:500, Swant, Switzerland), Glut1 (1:500, Abcam), PDGFR β (1:150 eBioscience) and incubated for 5 min in 4',6'-diaminodino-2-phenylindole (DAPI, 1:5000) before washing and mounting. The images were acquired on Zeiss LSM 710 inverted fluorescence confocal microscope (Carl Zeiss, Singapore). The same localization procedures were performed on hippocampal sections from the *Macaca fascicularis*. NeuN-immunostained neuronal profiles in the hippocampus and parvalbumin-stained Purkinje cells in cerebellum were counted and indicated as number per mm^2 (number of immunopositive neurons per square millimetre (number per mm^2) and number of immunopositive cells per millimetre (number per mm^2) in mean \pm s.e.m. Statistical significance was assessed using the Student's *t*-test.

Behaviour studies. For the Y-maze spontaneous-alternation test, mice were placed in the centre of a Y-shaped maze with three arms at a 120° angle from each other. The animal is allowed to freely explore all three arms. If the mouse has an intact spatial working memory, it would have a reduced tendency to enter the recently visited arm. The Topscan program (Cleversys, Reston, Virginia) was used to score the number of arm entries, and determine the number of alternations. A mouse is considered as having entered an arm when all four limbs are within the arm.

For the novel-object recognition test, the novel-object recognition test was carried out as described previously¹⁹. In brief, mice were trained with identical 'familiar' objects for 5 min and then assessed for short-term memory (STM) and long-term

memory (LTM), which took place 20 min and 24 h after training, respectively. The bouts of exploration for each object were scored using the Annostar program (Cleversys, Reston, Virginia). Preference score was calculated as (time spent with novel object – time spent with familiar object) ÷ (total time spent with both objects). Preferences for familiar object and novel object were defined as negative scores and scores approaching zero showed no preference for either object.

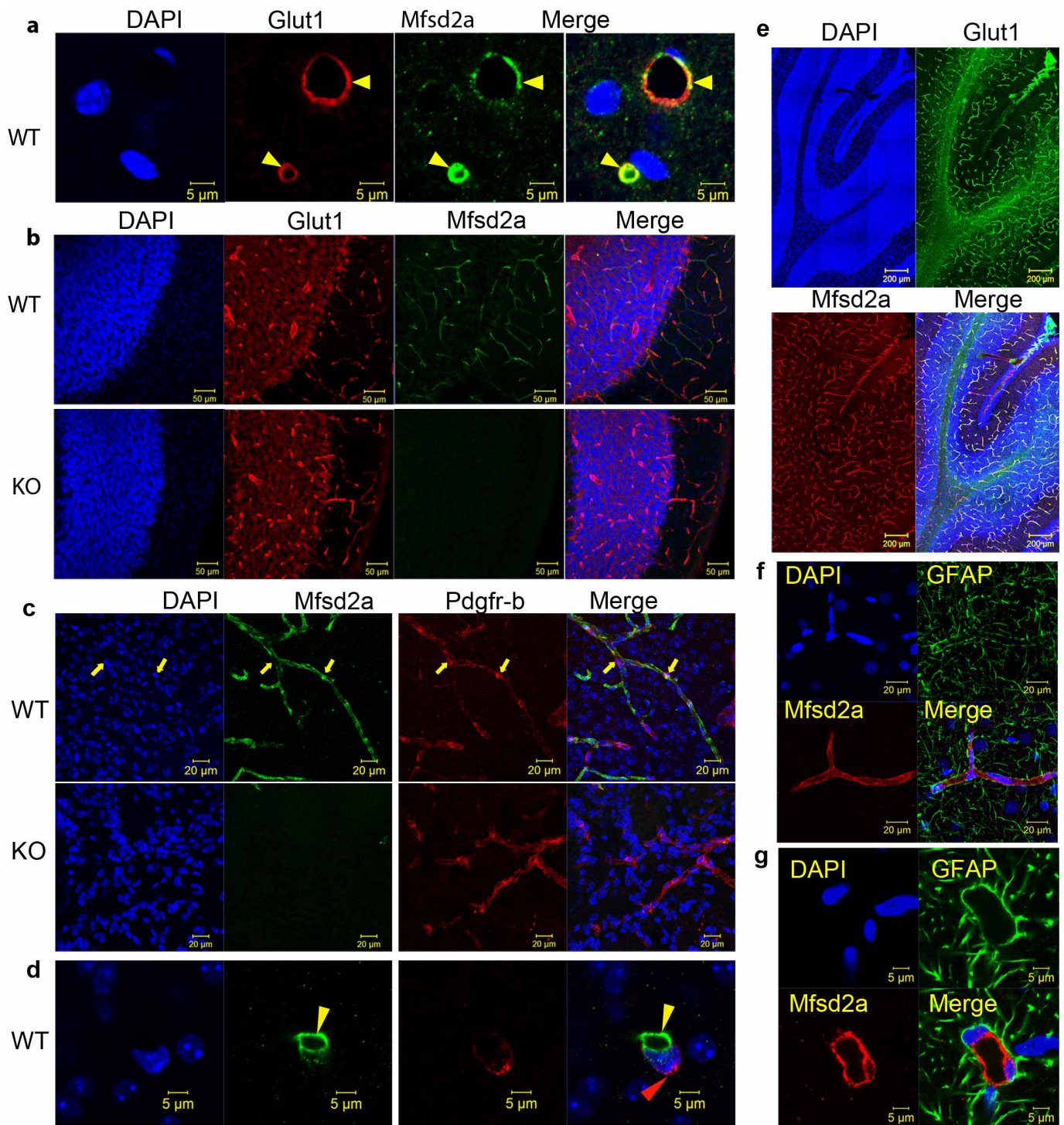
For anxiety tests, all behavioural apparatus were cleaned between each animal with surface disinfectant and 70% ethanol. In the zero maze, mice were placed in the closed arm and allowed 10 min for exploration. Behaviours, such as the time spent in, number of entries into, number of transitions between and latency to enter the open arms, were scored using the Annostar behaviour scoring program (Cleversys, Reston, Virginia, USA). In addition, exploratory behaviours, such as head dips and stretch attends, were recorded. At the start of the light–dark box test, mice were placed into the dark box measuring 20 × 40 × 16 cm, and allowed to move freely between the dark and light boxes for 10 min. Behavioural measures, such as time spent and horizontal activity in the light box, latency to enter the light box and the number of transitions between the two boxes, were recorded using the Versamax program (AccuScan Instruments, Columbus, Ohio, USA).

For open-field activity, mice were placed into the chamber for 60 min and the total distance travelled, number of rears, and time spent in the corners and centre were recorded using the Versamax program (AccuScan Instruments, Columbus, Ohio, USA).

Statistical analysis. Statistical differences of DHA and AA levels, and histological analyses between genotypes were calculated using an unpaired Student's *t*-test. Statistical analysis of LPC [¹⁴C]DHA, LPC [³H]palmitate, LPC [¹⁴C]oleate signals expressed as DPM between mock versus wild-type and mutants were calculated using two-way ANOVA; a *P* < 0.05 was considered to be significant.

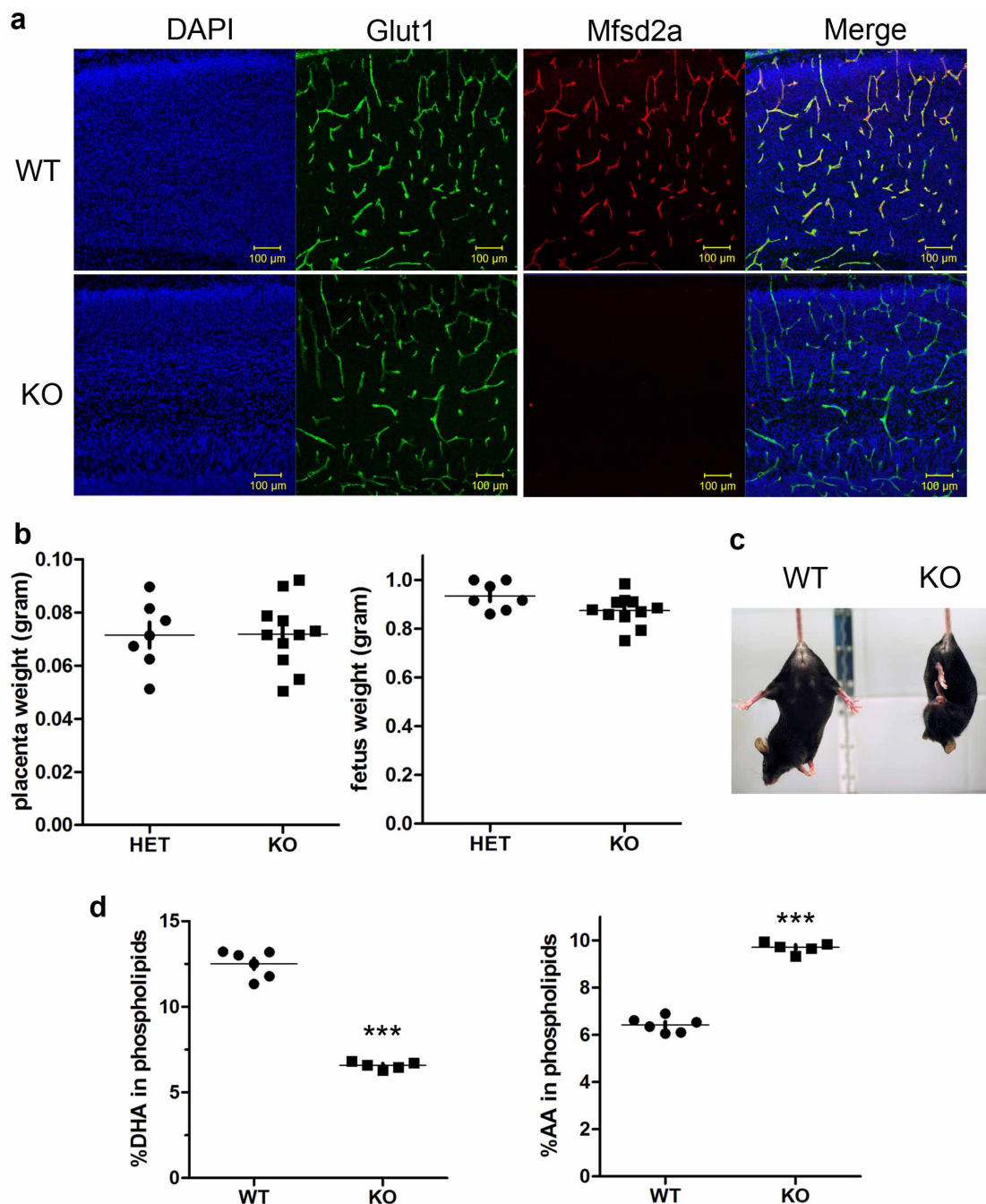
For the behavioural tests, genotype was the between subject factors. A one-way ANOVA was used to analyse the zero maze, light–dark box, open-field and Y-maze. A two-way ANOVA was used to analyse novel object test, with test day as the between subjects factor. Bonferroni corrected pair-wise comparisons were used as post-hoc tests. Data was expressed as mean ± s.e.m. and *P* < 0.05 was considered statistically significant.

19. Wong, P. *et al.* Pregnenolone rescues schizophrenia-like behavior in dopamine transporter knockout mice. *PLoS ONE* **7**, e51455 (2012).



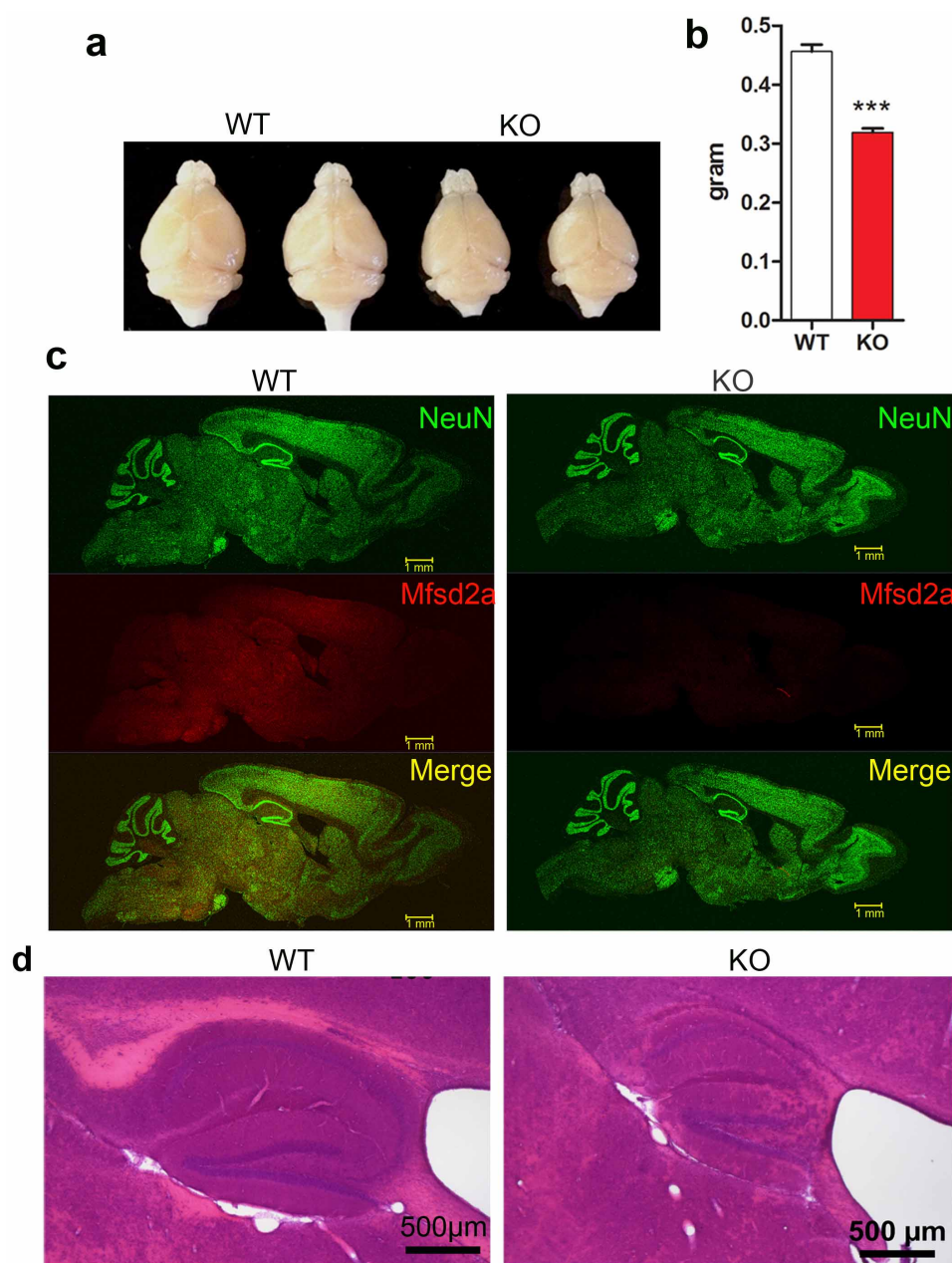
Extended Data Figure 1 | Mfsd2a is highly expressed in endothelium of micro-vessels in brain. **a**, Expression of Mfsd2a in endothelium is co-localized with glucose transporter Slc2a1 (Glut1). Arrowheads show endothelial cells in blood brain vessels. **b**, Mfsd2a is highly expressed in micro-vessels in brain, shown here are sections in dentate gyrus regions. **c**, Mfsd2a and the pericyte marker PDGFR β co-localize in brain microvasculature. **d**, Mfsd2a is not expressed in pericytes. Yellow and red arrowheads indicate endothelial cells and

a pericyte, respectively. Similar expression pattern of Mfsd2a is found in endothelium of micro-vessels in brain of monkey. **e**, Mfsd2a is highly expressed in micro-vessels and is co-localized with glucose transporter Slc2a1 (Glut1) in brain, shown here are sections in cerebellum of P4 monkey. **f**, **g**, Expression of Mfsd2a in endothelium in of brain micro-vessels. Shown is the hippocampal region. The marker for astrocytes is GFAP. KO, knockout; WT, wild type.



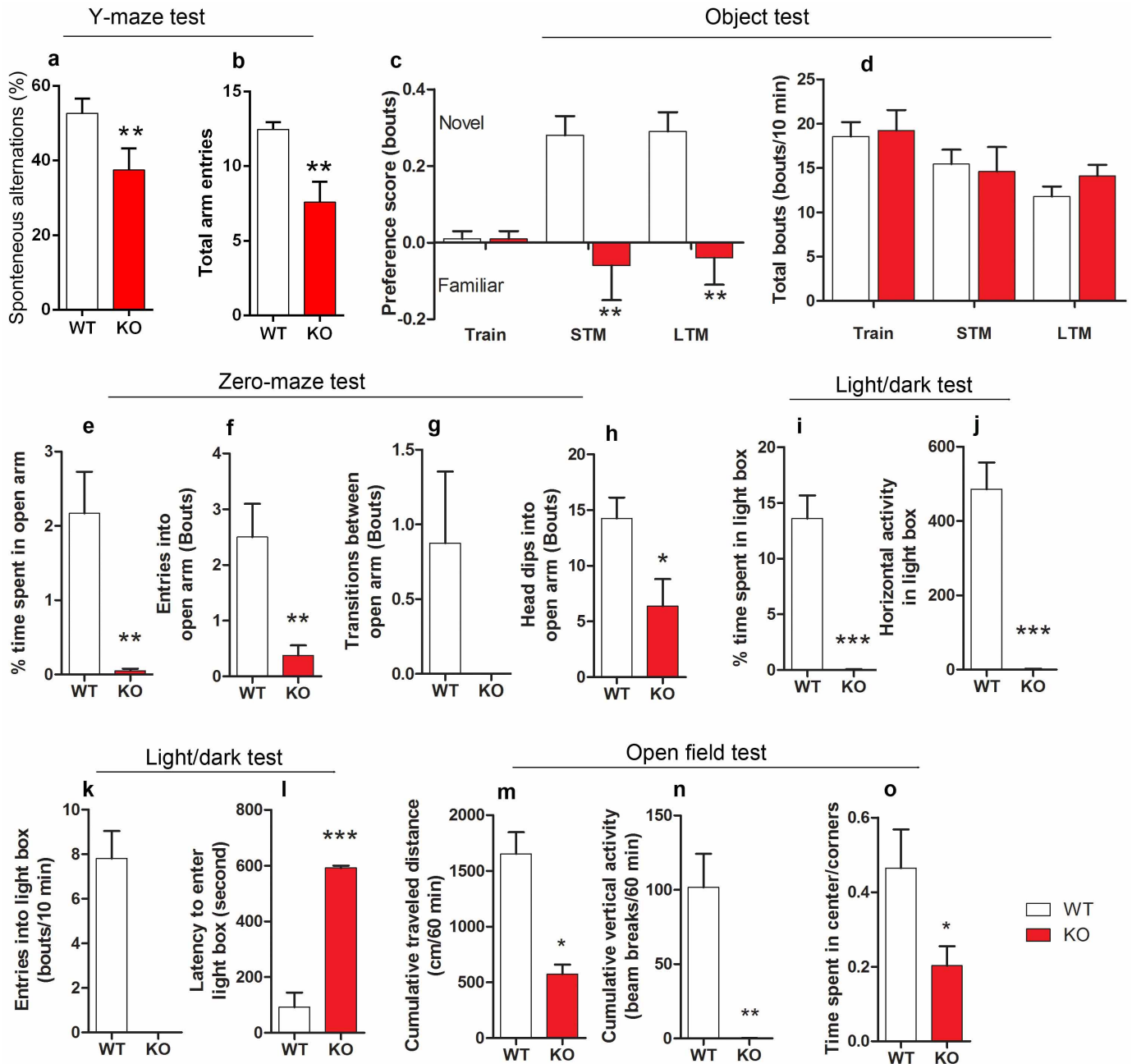
Extended Data Figure 2 | Localization of Mfsd2a at the blood–brain barrier of E15.5 fetus and lipid analysis. **a**, Mfsd2a (red) is highly expressed in micro-vessels and is co-localized with glucose transporter Slc2a1 (Glut1, green) in fetal brain. Scale bars 100 μ m. **b**, Placental and fetal weights. Placentas and fetuses from two HET pregnant mice (E18.5) crossed with a *Mfsd2a*-knockout male were collected and weighed. There were no significant differences in placental and fetal weight between HET ($n = 7$) and knockout ($n = 11$) mice.

c, Tail suspension was used to test for the presence of the paw-clasping phenotype of 10-week-old wild-type and knockout mice. **d**, Mass spectrometry measurement of phospholipids in the E18.5 fetal brain of wild-type ($n = 6$) and knockout ($n = 5$) mice showed that fetal brains of knockout mice had significantly reduced DHA levels, whereas AA levels were increased. *** $P < 0.001$ (see Source Data accompanying this figure for the full data set).



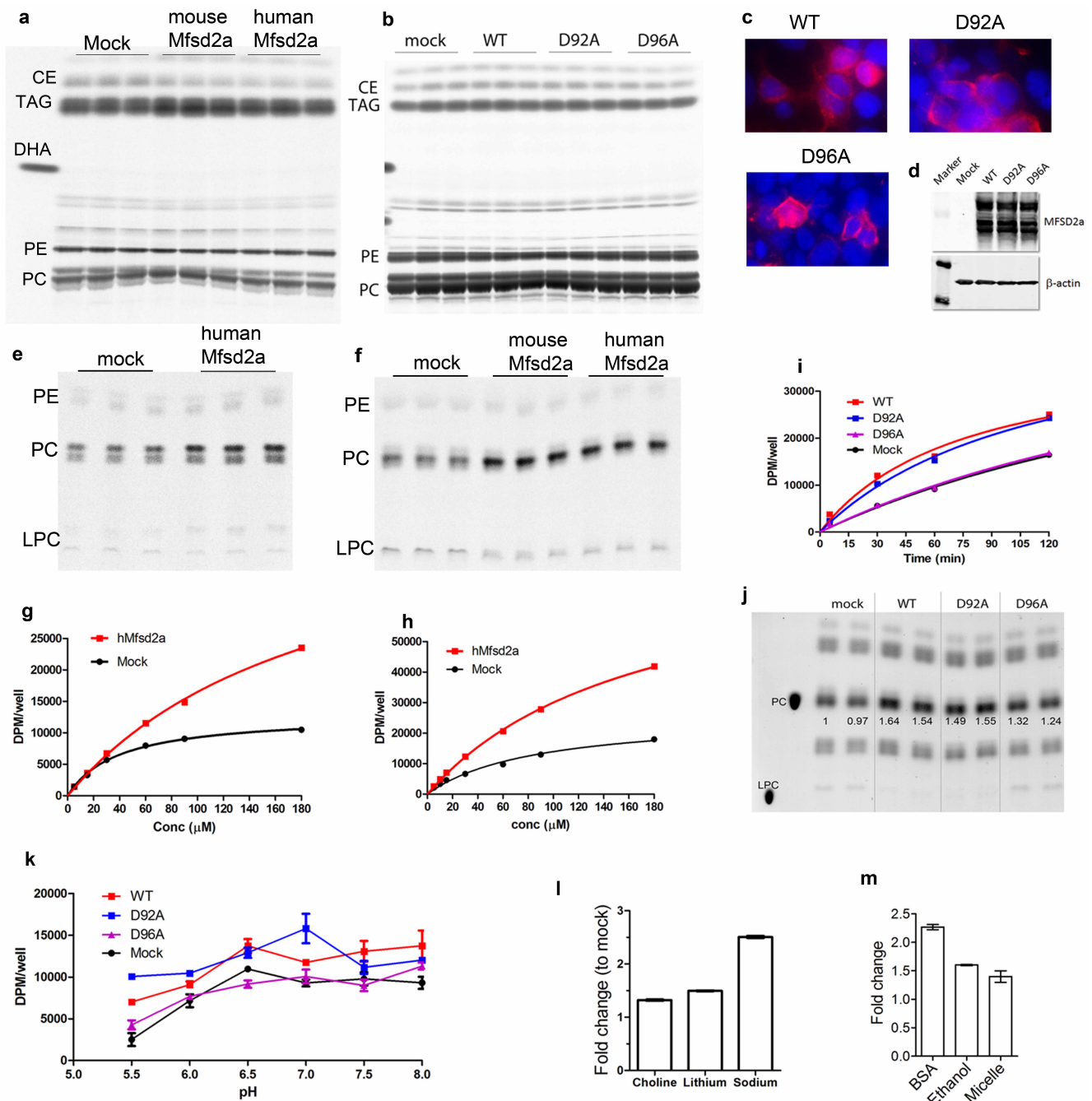
Extended Data Figure 3 | *Mfsd2a*-knockout mice exhibit microcephaly.
a, A representative image of brains of two 8-week-old wild-type and knockout littermates. **b**, The brain weight of knockout ($n = 4$) mice is significantly lower than that of wild-type ($n = 4$) littermates. **c**, Gross morphology of brains and sagittal sections of brains. Sagittal brain sections of 8-week-old wild-type and knockout mice were stained with NeuN to visualize neuronal cells and

Mfsd2a polyclonal antibody to visualize expression of *Mfsd2a*. *Mfsd2a* is shown to be widely expressed in brain. Scale bars, 1 mm. **d**, Haematoxylin and eosin (H&E) staining of hippocampus region of 8-week-old wild-type and knockout mice, indicating a smaller hippocampus in knockout mice. Scale bars, 500 μ m. *** $P < 0.001$. Data are expressed as mean \pm s.e.m.



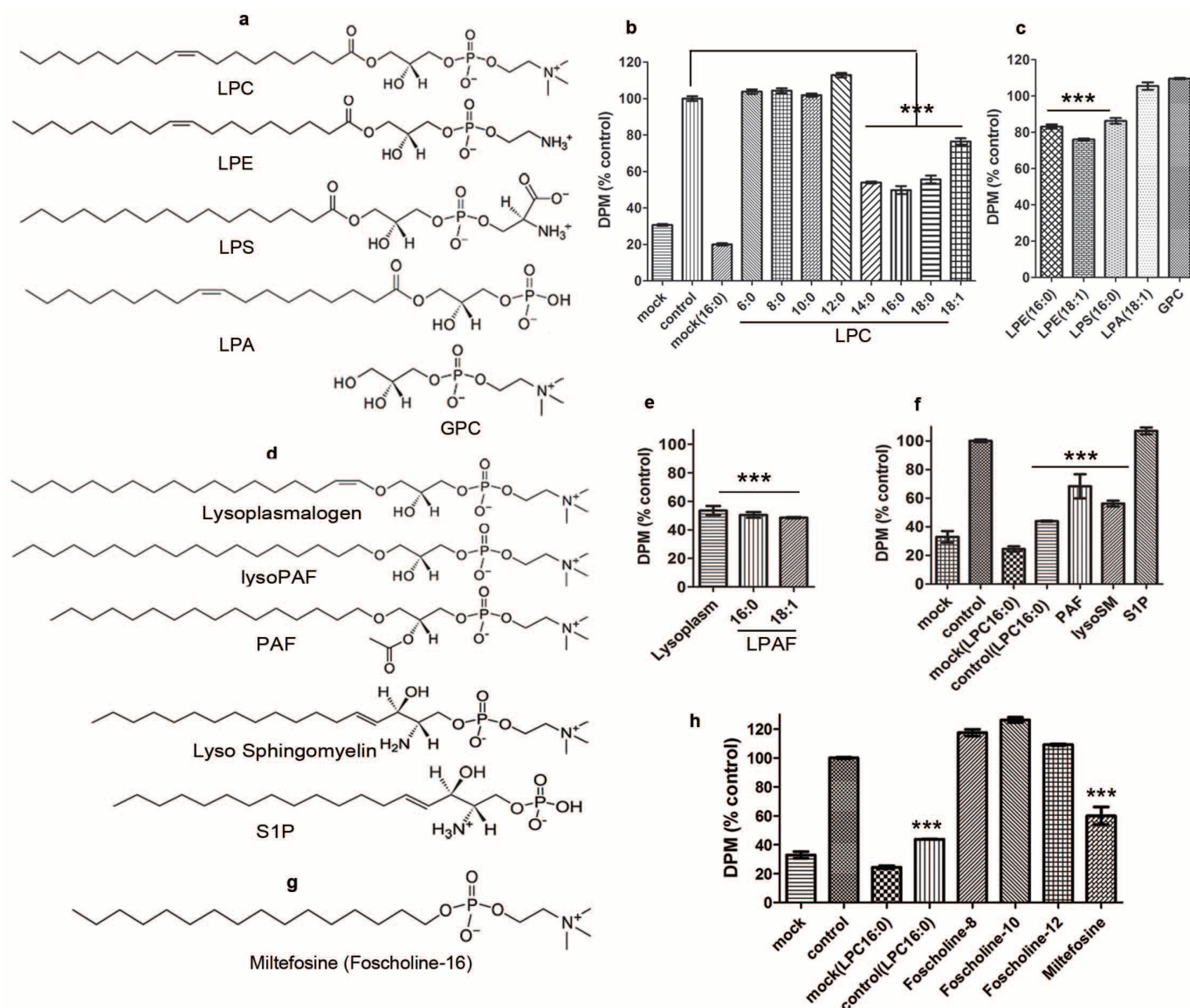
Extended Data Figure 4 | *Mfsd2a*-knockout mice exhibit deficits in learning and memory, and severe anxiety. a–d, The Y-maze test (a, b) and novel-object recognition test (c, d) were used to assess spatial learning, short-term memory (STM) and long-term memory (LTM) of the wild-type and knockout mice, respectively. Knockout mice exhibited significantly decreased total arm entries in a Y-maze test for spatial working memory. Knockout mice showed significantly reduced preferences for novel objects in novel-object recognition tests, indicative of defects in short-term memory and long-term memory, respectively. Train, the training period. e–l, Zero-maze tests (e–h) and light–dark box tests (i–l) were used to assess anxiety of the wild-type and knockout mice, respectively. Knockout mice showed decreased transitions and head dips into open arms during Zero-maze test for anxiety behaviours.

Knockout mice showed decreased entry into light box and increased latency to enter light box during light–dark box tests for anxiety. m–o, Open-field test for activity. Knockout mice showed reduced travel distance in the open-field test for locomotor activity. During the open-field test, knockout mice had no vertical activity indicative of motoric dysfunction, and decreased time spent in the centre, indicative of reduced exploration compared to wild-type mice. The increased time spent in the corners of the open field suggests that knockout mice were more anxious than wild-type mice, and are congruent with our results from the Zero-maze and light–dark box tests. Wild-type mice ($n = 11–13$) and knockout mice ($n = 8–10$). *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Data are expressed as mean \pm s.e.m.



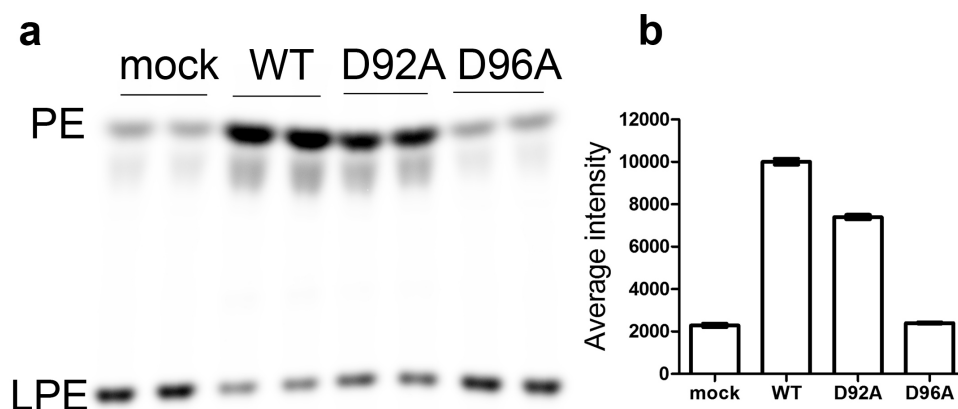
Extended Data Figure 5 | Human and mouse Mfsd2a transport LPCs but not unesterified fatty acids. **a**, Thin-layer chromatography (TLC) analysis of phospholipids and neutral lipids of HEK293 cells transfected with mouse Mfsd2a and human Mfsd2a after overnight incubation with 100 μ M [14 C]DHA. Std, unesterified [14 C]DHA. Experiments were repeated three times with triplicates. **b**, TLC analysis of phospholipids and neutral lipids of HEK293 cells transfected with mouse Mfsd2a and mutants after overnight incubation with 100 μ M [14 C]oleate. The TLC protocol used was described in the Methods. CE, cholesteryl ester; PC, phosphatidylcholine; PE, phosphatidylethanolamine; TAG, triglyceride. **c**, Localization of Mfsd2a, D92A and D96A at the plasma membrane (red). **d**, Western blot analysis of expression of Mfsd2a, D92A and D96A in HEK293 cells post 24 h transfection. **e**, Biological incorporation of radiolabelled LPC [14 C]DHA into PC. Cells expressing human (**e**, **f**) or mouse (**f**) Mfsd2a were incubated with LPC [14 C]DHA or 50 μ M LPC [14 C]oleate. Lipids were extracted from cells after 30 min incubation with LPC [14 C]DHA and 120 min incubation LPC [14 C]oleate and analysed using TLC method for resolving PC and LPC. Experiments were repeated two times with triplicate. **g**, **h**, Dose-dependent transport of LPC [14 C]DHA (**g**) and LPC [14 C]oleate

(**h**) by human Mfsd2a (hMfsd2a) and empty plasmid (mock) expressing HEK293. These experiments were performed in triplicate. **i**, Time-dependent transport of 50 μ M LPC [14 C]oleate. Mouse Mfsd2a (wild-type) and mutant constructs D92A and D96A were tested for uptake of radiolabelled LPCs at indicated times. This experiment was performed in triplicate. **j**, Increased net uptake of LPC ligand in cells expressing mouse Mfsd2a. TLC analysis of phospholipids of HEK293 cells transfected with mouse Mfsd2a and mutants after 1 h post incubation with 100 μ M unlabelled LPC oleate. Shown numbers are fold changes of PC levels relative to mock. Experiments were repeated three times with duplicates. **k**, Transport activity of mouse Mfsd2a (wild-type), D92A, D96A, and mock expressing cells was not significantly different at indicated pHs. This experiment was performed in triplicate. **l**, Activity of Mfsd2a is sodium- but not lithium-dependent. Data were expressed as fold change of Mfsd2a expressing cells compared to corresponding mock cells treated with the same conditions. This experiment was performed in triplicate. **m**, Transport activity of Mfsd2a is not BSA-dependent as LPC palmitate solubilized in either ethanol or micellar form was transported by Mfsd2a, albeit to a lower level than with BSA. This experiment was performed in triplicate. Data are expressed as mean \pm s.e.m.



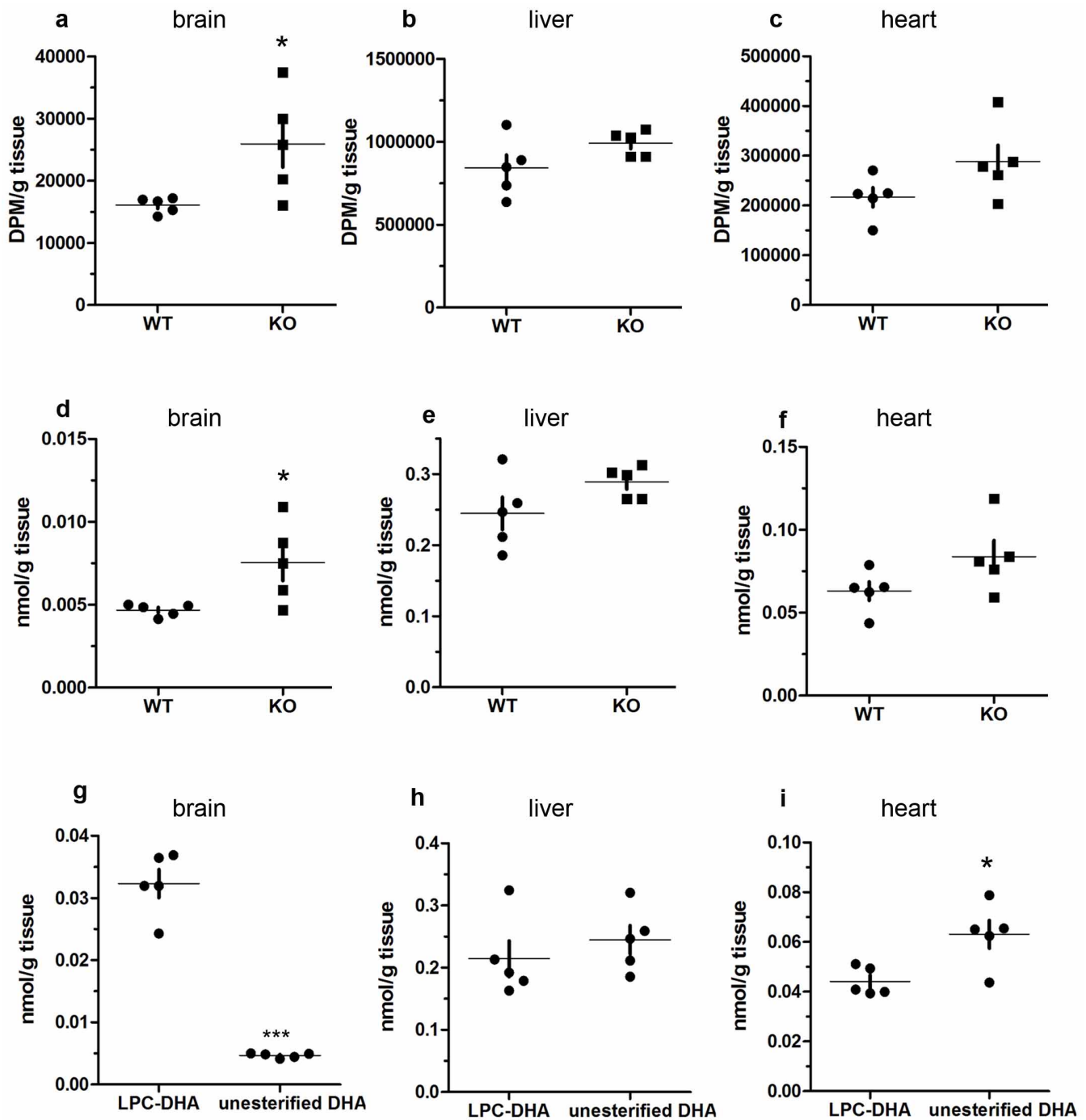
Extended Data Figure 6 | Competition assay to determine the ligand structures of Mfsd2a. All competition assays were performed using 25 μ M LPC [3 H]palmitate as ligand with or without tenfold molar excess (250 μ M) of the indicated competitors. **a**, The structures of the lipid competitors used in **b** and **c**. **b**, Competition assays with indicated acyl chain LPCs. **c**, Competition assay with indicated headgroups. Assays were stopped after 30 min of incubation. Competitive activity was expressed as percent to control (activity of Mfsd2a without competitor). 6:0, hexanoate; 8:0, octanoate; 10:0, docanoate; 12:0, laurate; 14:0, myristate; 16:0, palmitate; 18:0, stearate; 18:1, oleate; GPC, alpha-glycerolphosphocholine; LPA, lysophosphatidic acid; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPS, lysophosphatidylserine. **d**, Representative structures of bioactive lipid competitors used in **e** and **f**. **e**, Competition assay with lysophospholipid forms of plasmalogens and platelet activating factor (PAF). This experiment was

performed together with **b**, so that the control and mock shown in **b** can be used for reference. **f**, PAF and lysosphingomyelin (lysoSM) also showed strong competition, whereas sphingosine 1-phosphate (S1P) did not compete for LPC [3 H]16:0 uptake. Competitive activity was expressed as percent to control (activity of Mfsd2a without competitor). **g**, Representative structure of non-biological lysophospholipid analogues foscholine-16 (Miltefosine) with an alkyl chain of 16 carbons. **h**, Competition assays of indicated foscholines with LPC [3 H]palmitate. Assays were stopped after 15 min of incubation. Competitive activity was expressed as percent to control (activity of Mfsd2a without competitor). Foscholine with alkyl chain length of 8 (Fos-8), 10 (Fos-10) and 12 (Fos-12) carbons did not compete, whereas foscholine with an alkyl chain length of 16 carbons (Miltefosine) showed strong competition with LPC [3 H]16:0. These experiments were repeated two times with triplicates. *** $P < 0.001$. Data are expressed as mean \pm s.e.m.



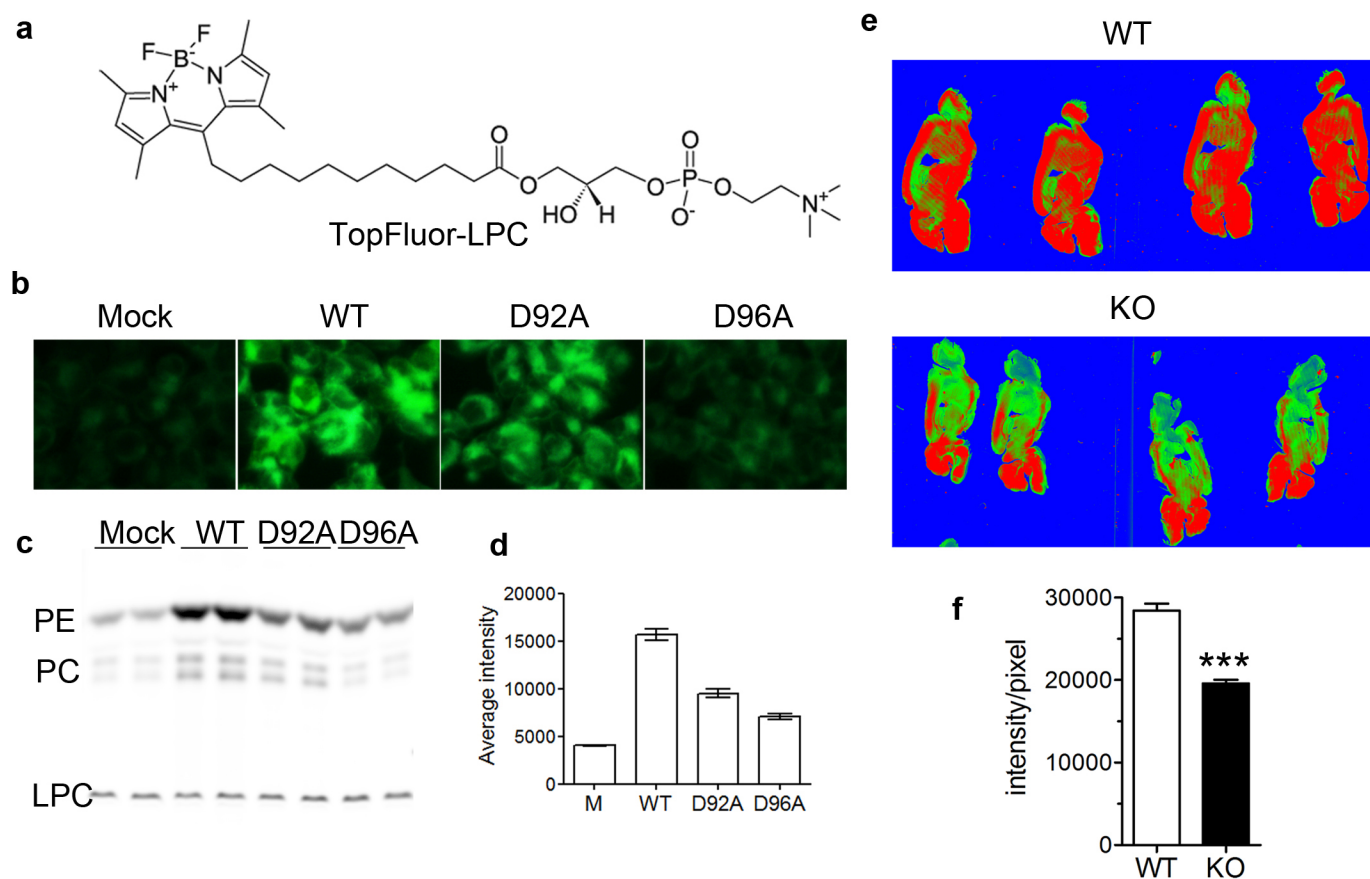
Extended Data Figure 7 | Mfsd2a transports TopFluor LPE. Thin-layer chromatography (TLC) analysis of phospholipids of HEK293 cells transfected with mouse Mfsd2a and mutants after 30 min post incubation with 25 μ M TopFluor LPE. **a**, TLC analysis of phospholipids. **b**, Quantification of intensity

of PE band from TLC plate. PE, phosphatidylethanolamine; LPE, lysophosphatidylethanolamine. Experiments were repeated two times with duplicates. Data are expressed as mean \pm s.e.m.



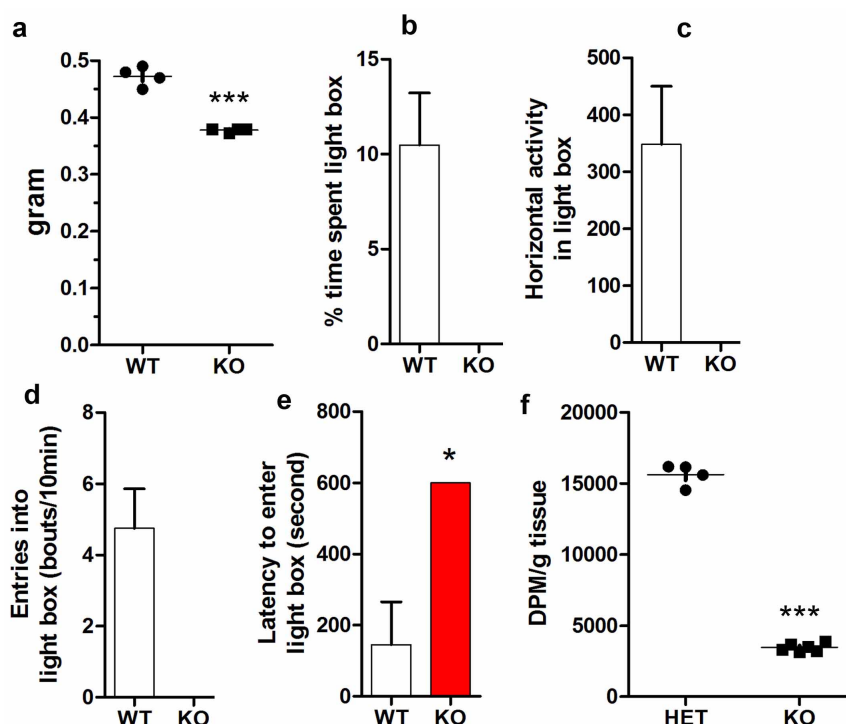
Extended Data Figure 8 | Brain uptake of unesterified [^{14}C]DHA was not reduced in *Mfsd2a*-knockout mice. Male mice aged 7 weeks old were intravenously injected with 1 mmol of [^{14}C]DHA-BSA complex. Brain, liver and heart were collected 2 h post injection for lipid extraction and DPM quantified using scintillation counting. **a–c**, Uptake of unesterified [^{14}C]DHA in the wild-type and knockout brain, heart and liver were expressed as DPM per g. **d–f**, Level of uptake of DHA in **a–c** was converted into nmol per g.

g–i, A comparison between the absolute amount of DHA uptake in the form of LPC DHA (converted from Fig. 4 into nmol per g in 2 h) and unesterified DHA (taken from **d–f** above) in brain, heart and liver of wild-type mice. The same amounts of LPC DHA and DHA were injected in mice. The amount of LPC DHA uptake was far greater than unesterified DHA uptake by wild-type brain. *** $P < 0.0001$, * $P < 0.05$. Data are expressed as mean \pm s.e.m. (wild-type, $n = 5$; knockout, $n = 5$).



Extended Data Figure 9 | Brain uptake of TopFluor LPC was reduced in *Mfsd2a*-knockout mice. This experiment was carried out as described for NBD LPC in Fig. 4. **a**, structure of TopFluor LPC. **b**, HEK293 cells expressing wild-type *Mfsd2a* showed significantly enhanced uptake activity to TopFluor LPC compared with mock (empty plasmid), D92A and D96A mutant expressing cells. **c**, TLC analysis showed that TopFluor LPC was bio-incorporated into PE. Experiments were repeated two times with

duplicates. **d**, Quantification of PE band from the TLC plate shown in **c**. **e**, Brain uptake of TopFluor LPC was decreased in knockout mice. Male mice (wild-type, $n = 3$; knockout, $n = 3$) aged 7 weeks old were intravenously injected with 300 μg TopFluor LPC-BSA complex. **f**, Fluorescence from 10 brain sections of wild-type and knockout mice was quantified and expressed as fluorescence intensity per pixel. *** $P < 0.001$. Data are expressed as mean \pm s.e.m.



Extended Data Figure 10 | Dietary DHA supplementation failed to rescue *Mfsd2a*-knockout phenotypes. Heterozygous female mice were gavaged with 100 μ l DHA oil (containing 26% DHA triglyceride and 6% EPA, total omega-3 is 35%) every 2 days for 2 weeks before conception in crosses with heterozygous males. During gestation, pregnant mice were continued on gavages of DHA every 2 days. Gavages of mothers continued during breastfeeding and pups were weaned onto normal diet at 3 weeks of age and gavaged every 2 days with DHA for 8 weeks. **a**, Brain weight of adult wild-type ($n = 4$) and knockout ($n = 4$) mice aged 8 weeks after treatment with dietary DHA oil. Knockout mice brains were still significantly smaller. **b–e**, DHA treatment of knockout mice did not reduce the strong anxiety phenotype as determined using the light–dark box test. To investigate why dietary DHA failed to rescue knockout phenotypes, we tested the hypothesis that uptake of

maternally derived DHA (in this case the DHA delivered to the mother via gavage) might not get into the brain of knockout mice during brain development. To test this possibility, pregnant HET mothers intercrossed with knockout fathers were gavaged at E17.5–E19.5 with [14 C]DHA and uptake into fetal brains was quantified. Note that heterozygous mice do not exhibit detectable phenotypes, have similar DHA to wild type (not shown) and are thus similar to wild type. The rationale for using HET and knockout intercrosses was to increase the yield of knockout mice in this study. The data shown in **f** indicate that brains of knockout mice exhibited an 80% reduction in the uptake of [14 C]DHA relative to HET mice with the same mothers ($n = 4$ wild-type, $n = 6$ knockout). Therefore, *Mfsd2a* expressed during fetal development is important for DHA transport into brain. * $P < 0.05$, *** $P < 0.001$. Data are expressed as mean \pm s.e.m.

Mfsd2a is critical for the formation and function of the blood–brain barrier

Ayal Ben-Zvi¹, Baptiste Lacoste¹, Esther Kur¹, Benjamin J. Andreone¹, Yoav Mayshar², Han Yan¹ & Chenghua Gu¹

The central nervous system (CNS) requires a tightly controlled environment free of toxins and pathogens to provide the proper chemical composition for neural function. This environment is maintained by the ‘blood–brain barrier’ (BBB), which is composed of blood vessels whose endothelial cells display specialized tight junctions and extremely low rates of transcellular vesicular transport (transcytosis)^{1–3}. In concert with pericytes and astrocytes, this unique brain endothelial physiological barrier seals the CNS and controls substance influx and efflux^{4–6}. Although BBB breakdown has recently been associated with initiation and perpetuation of various neurological disorders, an intact BBB is a major obstacle for drug delivery to the CNS^{7–10}. A limited understanding of the molecular mechanisms that control BBB formation has hindered our ability to manipulate the BBB in disease and therapy. Here we identify mechanisms governing the establishment of a functional BBB. First, using a novel tracer-injection method for embryos, we demonstrate spatiotemporal developmental profiles of BBB functionality and find that the mouse BBB becomes functional at embryonic day 15.5 (E15.5). We then screen for BBB-specific genes expressed during BBB formation, and find that major facilitator super family domain containing 2a (*Mfsd2a*) is selectively expressed in BBB-containing blood vessels in the CNS. Genetic ablation of *Mfsd2a* results in a leaky BBB from embryonic stages through to adulthood, but the normal patterning of vascular networks is maintained. Electron microscopy examination reveals a dramatic increase in CNS-endothelial-cell vesicular transcytosis in *Mfsd2a*^{−/−} mice, without obvious tight-junction defects. Finally we show that *Mfsd2a* endothelial expression is regulated by pericytes to facilitate BBB integrity. These findings identify *Mfsd2a* as a key regulator of BBB function that may act by suppressing transcytosis in CNS endothelial cells. Furthermore, our findings may aid in efforts to develop therapeutic approaches for CNS drug delivery.

Two unique features of the CNS endothelium determine BBB integrity (Extended Data Fig. 1)². One is specialized tight junctions between a single endothelial cell layer lining the CNS capillaries, which form the physical seal between the blood and brain parenchyma². In addition, CNS endothelial cells have lower rates of transcytosis than endothelial cells in other organs³. Peripheral endothelial cells display active vesicle trafficking to deliver nutrients to peripheral tissues, whereas CNS endothelial cells express transporters to selectively traffic nutrients across the BBB^{13,11}. However, it is not clear when and how these properties are acquired. Furthermore, the molecular mechanisms that give rise to the unique properties of the CNS endothelium have not been identified. Although recent studies revealed molecular pathways involved in the development of the embryonic BBB^{12–19}, disruption of some of these genes affect vascular network development, making it difficult to determine whether barrier defects are primary or secondary to a broader vascular effect.

We aimed to first identify the developmental time-point when the BBB gains functional integrity, and then use that time window to profile BBB-specific genes when the BBB is actively forming, to maximize the chance of identifying key regulators. The prevailing view has been that the embryonic and perinatal BBB are not yet functional¹. However, previous

embryonic BBB functionality studies were primarily performed by transcardiac tracer perfusion, which may dramatically affect blood pressure, cause bursting of CNS capillaries, and artificially produce leakiness phenotypes^{1,20}. To circumvent these obstacles, we developed a method to assess BBB integrity during mouse development, in which a small volume of tracer is injected into embryonic liver to minimize changes in blood pressure (Fig. 1a, see Supplementary Information for the method).

Using this method, we identified the timing of BBB formation in the developing mouse brain and observed a spatial and temporal pattern of ‘functional-barrier genesis’ (Fig. 1b). We found that in E13.5 cortex a 10-kDa dextran tracer leaked out of capillaries and was taken up by non-vascular brain parenchyma cells (Fig. 1b, top panel). At E14.5, the tracer was primarily restricted to capillaries, but tracer was still detected outside vessels (Fig. 1b, middle panel). In contrast, at E15.5, the tracer was confined to vessels with no detectable signal in the surrounding brain parenchyma, similar to the mature BBB (Fig. 1b, bottom panel). The development of BBB functionality differed across brain regions (Supplementary Information and Extended Data Fig. 2). These data demonstrate that following vessel ingression into the neural tube, the BBB gradually becomes functional as early as E15.5.

Based on the temporal profile of BBB formation, we compared expression profiles of BBB (cortex) and non-BBB (lung) endothelium at E13.5, using an Affymetrix array (Supplementary Information), and identified transcripts with significantly higher representation in cortical than lung endothelium (Fig. 2). These transcripts included transporters, transcription factors, and secreted and transmembrane proteins (Fig. 2c). We were particularly interested in transmembrane proteins, owing to their potential involvement in cell–cell interactions that regulate BBB formation.

One of the genes identified, *Mfsd2a*, had 78.8 times higher expression in cortical endothelium than in lung endothelium (Fig. 3a). *In situ* hybridization showed prominent *Mfsd2a* mRNA expression in CNS vasculature but no detectable signal in vasculature outside the CNS, such as in lung or liver (Fig. 3b). Moreover, both *Mfsd2a* mRNA and *Mfsd2a* protein were absent in the choroid plexus vasculature, which is part of the CNS but does not possess a BBB¹ (Fig. 3c, d, g). *Mfsd2a* expression in CNS vasculature was observed at embryonic stages (E15.5), postnatal days 2 and 5 (P2 and P5) and in adults (P90) (Fig. 3b–e and Extended Data Fig. 3). Finally, *Mfsd2a* protein, which is absent in the *Mfsd2a*^{−/−} mice (Fig. 3e)²¹, was specifically expressed in claudin-5-positive CNS endothelial cells but not in neighbouring parenchyma cells (neurons or glia) or adjacent Pdgfrβ-positive pericytes (Fig. 3f). Previously, *Mfsd2a* was reported to be a transmembrane protein expressed in the placenta and testis, which have highly restrictive barrier properties²². Together with our demonstration of *Mfsd2a*-specific expression in BBB-containing endothelial cells, this suggests that *Mfsd2a* may have a role in BBB formation and/or function.

To test this hypothesis, we examined BBB integrity in *Mfsd2a*^{−/−} mice. Using our embryonic injection method, 10-kDa dextran was injected into *Mfsd2a*^{−/−} and wild-type littermates at E15.5. As expected, dextran was confined within vessels of control embryos. In contrast, dextran

¹Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Genetics, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

leaked outside the vessels in *Mfsd2a*^{-/-} embryonic brains and was found in the cortical parenchyma (Fig. 4a) and individual parenchyma cells (quantified as tracer-positive parenchyma cells per unit area of the developing lateral cortical plate; Fig. 4b). Furthermore, using imaging and spectrophotometric quantification methods⁵, we found that the leaky phenotype persisted in early postnatal (Extended Data Fig. 4) and adult (Fig. 4c) *Mfsd2a*^{-/-} mice. Because the sequence of *Mfsd2a* has similarities to the major facilitator superfamily of transporters, and *Mfsd2a* facilitates the transport of tunicamycin in cancer cell lines²³, we injected two non-carbohydrate-based tracers of different sizes to rule out the possibility that dextran leakiness is due to interactions with *Mfsd2a*. Sulfo-NHS-biotin (~550 Da) and horseradish peroxidase (HRP; ~44 kDa) tracers exhibited the leaky phenotype in *Mfsd2a*^{-/-} mice (Extended Data Fig. 4a, b). Moreover, a larger molecular weight tracer, 70-kDa dextran, also displayed leakiness in *Mfsd2a*^{-/-} mice (Extended Data Fig. 4d). In contrast to severe barrier leakage defects (Fig. 4a–c and Extended Data Fig. 4), brain vascular patterning was similar between *Mfsd2a*^{-/-} mice and littermate controls. No abnormalities were identified in capillary density, capillary diameter or vascular branching (Fig. 4d and Extended Data Fig. 5a), in embryonic (E15.5), postnatal (P4), and adult (P70) brains of *Mfsd2a*^{-/-} mice. Moreover, we found no abnormalities in cortical arterial distribution in adult *Mfsd2a*^{-/-} mice (Extended Data

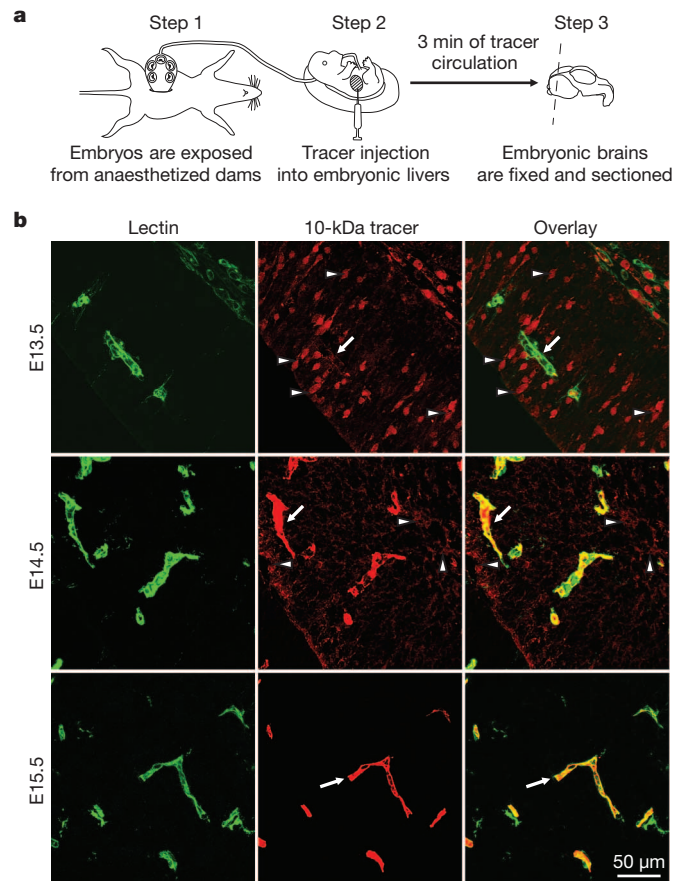


Figure 1 | A novel tracer-injection method reveals a temporal profile of functional BBB formation in the embryonic cortex. **a**, *In utero* embryonic liver tracer injection method; fenestrated liver vasculature enabled rapid tracer uptake into the embryonic circulation. **b**, Dextran-tracer injection revealed a temporal profile of functional cortical BBB formation. Representative images of dorsal cortical plates from injected embryos after capillary labelling with lectin (green, lectin; red, 10-kDa tracer). Top panel (E13.5), tracer leaked out of capillaries and was subsequently taken up by non-vascular parenchyma cells (arrowheads), with little tracer left inside capillaries (arrow). Middle panel (E14.5), tracer was primarily restricted to capillaries (arrow), with diffused tracer detectable in the parenchyma (arrowheads). Bottom panel (E15.5), tracer was confined to capillaries (arrow). *n* = 6 embryos (3 litters per age).

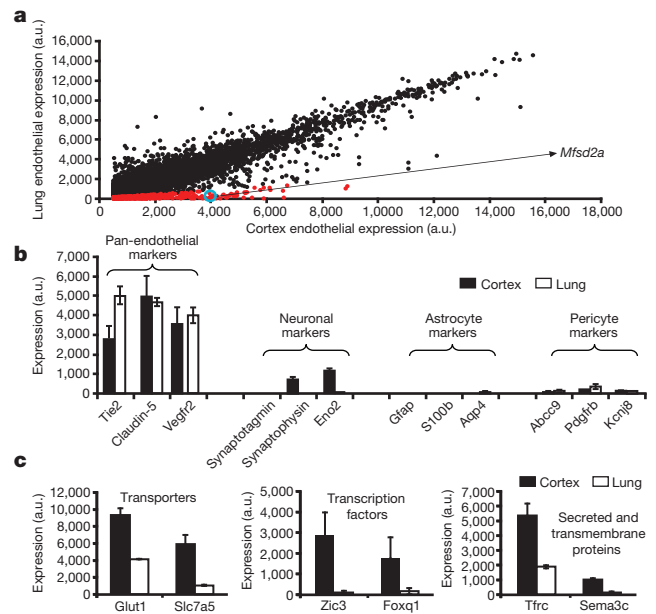


Figure 2 | Expression profiling identifies genes involved in BBB formation.

a, Dot-plot representation of Affymetrix GeneChip data showing transcriptional profile of cortical (BBB) and lung (non-BBB) endothelial cells isolated at the critical barrier-genesis period (E13.5). Dots reflect average expression of a probe in the cortex (x axis) and lung (y axis). Cortex expression values above 500 arbitrary expression units (a.u.) are presented. Red dots indicate a fivefold higher expression in the cortex. *Mfsd2a* value is circled in blue. **b**, Pan-endothelial markers were highly represented, whereas pericyte, astrocyte, and neuronal markers were detected at extremely low levels in both cortex and lung samples. **c**, Barrier-genesis specific transporters, transcription factors, and secreted and transmembrane proteins were significantly enriched in the cortical endothelial cells. All data are mean \pm s.d. *n* = 4 litters (4 biological replicates).

Fig. 5b). Therefore, *Mfsd2a* is specifically required for proper formation of a functional BBB but not for CNS vascular morphogenesis *in vivo*. This result, together with the temporal difference between cortical vascular ingression (E10–E11) and cortical barrier-genesis (E13.5–E15.5), demonstrates that vascular morphogenesis and barrier genesis are distinct processes.

We next addressed whether *Mfsd2a* regulates endothelial tight-junction formation, transcytosis, or both. We examined these properties by electron microscopy in embryonic brains and P90 mice following intravenous HRP injection². Electron microscopy failed to reveal any apparent abnormalities in the ultrastructure of endothelial tight junctions (Fig. 5a). At E17.5, tight junctions in control and *Mfsd2a*^{-/-} littermates appeared normal, with electron-dense linear structures showing 'kissing points' where adjacent membranes are tightly apposed (Fig. 5a). In electron micrographs of cerebral cortex in HRP-injected adults, peroxidase activity was revealed by an electron-dense reaction product that filled the vessel lumen. In both control and *Mfsd2a*^{-/-} mice, HRP penetrated the intercellular spaces between neighbouring endothelial cells only for short distances. HRP was stopped at the tight junction, creating a boundary between HRP-positive and HRP-negative regions without leakage through tight junctions (Fig. 5a). In contrast, CNS endothelium of *Mfsd2a*^{-/-} mice displayed a dramatic increase in the number of vesicles, including luminal and abluminal plasma membrane-connected vesicles and free cytoplasmic vesicles, which may indicate an increased rate of transcytosis (Fig. 5b). Specifically, pinocytotic events were evidenced by type II lumen-connected vesicles pinching from the luminal plasma membrane. Greater than twofold increases in vesicle number in *Mfsd2a*^{-/-} mice compared to control littermates were observed in different locations along the transcytotic pathway (Fig. 5 and Extended Data Table 2). Furthermore, the HRP reaction product in adult mice was observed in vesicles

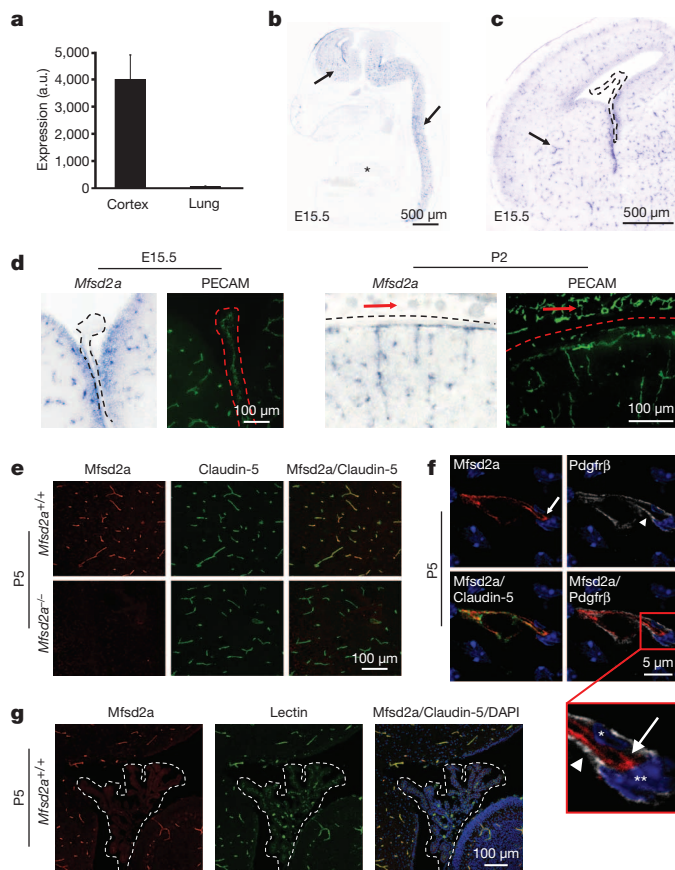


Figure 3 | *Mfsd2a* is selectively expressed in BBB-containing CNS vasculature. **a**, At E13.5, *Mfsd2a* expression in cortical endothelium was ~80-fold higher than in lung endothelium (microarray analysis, mean \pm s.d.). **b–d**, Specific *Mfsd2a* expression in BBB-containing CNS vasculature (blue, *Mfsd2a* in situ hybridization; green, vessel staining (PECAM) adjacent sections). **b**, *Mfsd2a* expression at E15.5 in CNS vasculature (sagittal view of brain and spinal cord, arrows), but not in non-CNS vasculature (asterisk). **c**, *Mfsd2a* expression at E15.5 in BBB vasculature (cortex coronal view, for example, striatum, arrow), but not in non-BBB CNS vasculature (choroid plexus, dashed line). **d**, High-magnification coronal view of *Mfsd2a* expression in BBB-containing CNS vasculature but not in vasculature of the choroid plexus (left, dashed line), or outer meninges or skin (right, red arrows). **e–g**, Immunohistochemical staining of *Mfsd2a* protein shows specific expression in CNS endothelial cells (red, *Mfsd2a*; green, claudin-5 or lectin (endothelium); blue, DAPI (nuclei); grey, Pdgfr β (pericytes)). **e**, *Mfsd2a* expression in the brain vasculature of wild-type mice (top panel), but not of *Mfsd2a*^{-/-} mice (bottom panel). **f**, *Mfsd2a* expression only in claudin-5-positive endothelial cells (arrow; endothelial nucleus is indicated by an asterisk) but not in adjacent pericytes (arrowhead; pericyte nucleus is indicated by a double asterisk). **g**, Lack of *Mfsd2a* expression in choroid plexus vasculature (fourth ventricle coronal view, dashed line), as opposed to the prominent *Mfsd2a* expression in cerebellar vasculature. $n = 3$ embryos (3 litters per age).

that the BBB leakiness observed in *Mfsd2a*^{-/-} mice was not caused by opening of tight junctions, but rather by increased transcellular trafficking across the endothelial cytoplasm.

Studies using pericyte-deficient genetic mouse models have shown that pericytes can also regulate BBB integrity. These mice had increased vesicle trafficking without obvious junction defects^{4,5}, similar to our observations in *Mfsd2a*^{-/-} mice. We therefore examined the possibilities that *Mfsd2a* may regulate CNS endothelial transcytosis by modulating pericyte function or that the effect of pericytes on endothelial transcytosis is mediated by *Mfsd2a*. First, pericyte coverage, attachment to the capillary wall, and pericyte ultrastructure and positioning relative to endothelial cells were normal in *Mfsd2a*^{-/-} mice (Extended Data Fig. 6). These data, together with the lack of *Mfsd2a* expression in pericytes, suggest that the increased transcytosis observed in *Mfsd2a*^{-/-} endothelial cells is not secondary to pericyte abnormalities. Second, a genetic reduction in pericyte coverage can influence endothelial gene expression

invaginated from the luminal membrane and exocytosed at the abluminal plasma membrane only in *Mfsd2a*^{-/-} mice (Fig. 5d), suggesting that HRP was subject to transcytosis in these animals but not in wild-type littermates (Extended Data Table 2). Together, these findings suggest

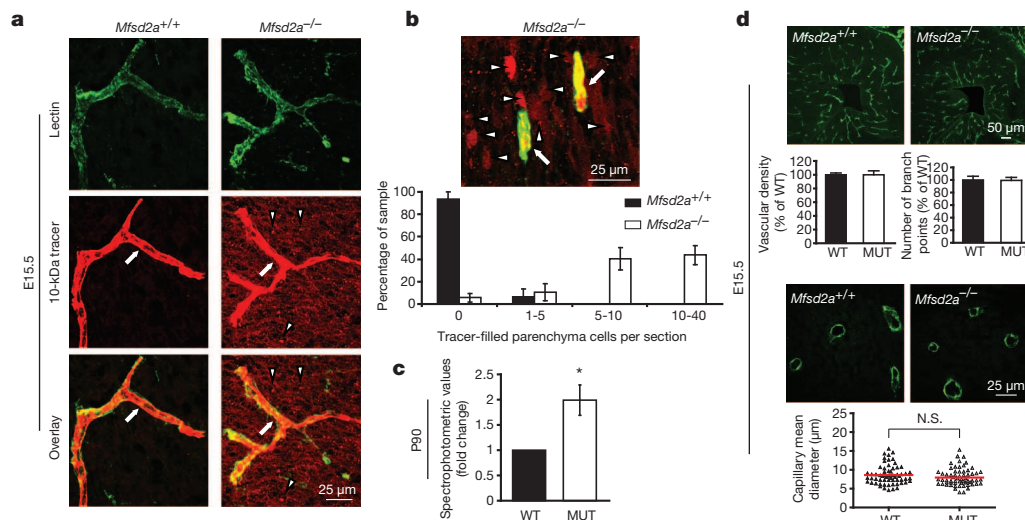


Figure 4 | *Mfsd2a* is required for the establishment of a functional BBB but not for CNS vascular patterning in vivo. **a**, **b**, Dextran-tracer (10 kDa) injections at E15.5 revealed a defective BBB in mice lacking *Mfsd2a*. **a**, The tracer was confined to the capillaries (arrow) in wild-type littermates, whereas *Mfsd2a*^{-/-} embryos showed large amounts of tracer leakage in the brain parenchyma (arrowheads). **b**, Capillaries (arrows) surrounded by tracer-filled brain parenchyma cells (arrowheads) in *Mfsd2a*^{-/-} cortex. Quantification of tracer-filled parenchyma cells in control versus *Mfsd2a*^{-/-} cortical plates (bottom panel, $n = 7$ embryos per genotype). **c**, Spectrophotometric

quantification of 10-kDa dextran-tracer from cortical extracts of P90 mice, 16 h post intravenous injection, indicating that BBB leakiness in *Mfsd2a*^{-/-} mice persists into adulthood ($n = 3$ mice per genotype). **d**, *Mfsd2a*^{-/-} mice exhibit normal vascular patterning. No abnormalities were found in cortical vascular density, branching and capillary diameter (E15.5; green, PECAM). Quantification of wild-type and *Mfsd2a*^{-/-} samples ($n = 4$ embryos per genotype). All data are mean \pm s.e.m. MUT, mutant; N.S., not significant; WT, wild type. * $P < 0.05$ (Mann–Whitney U -test).

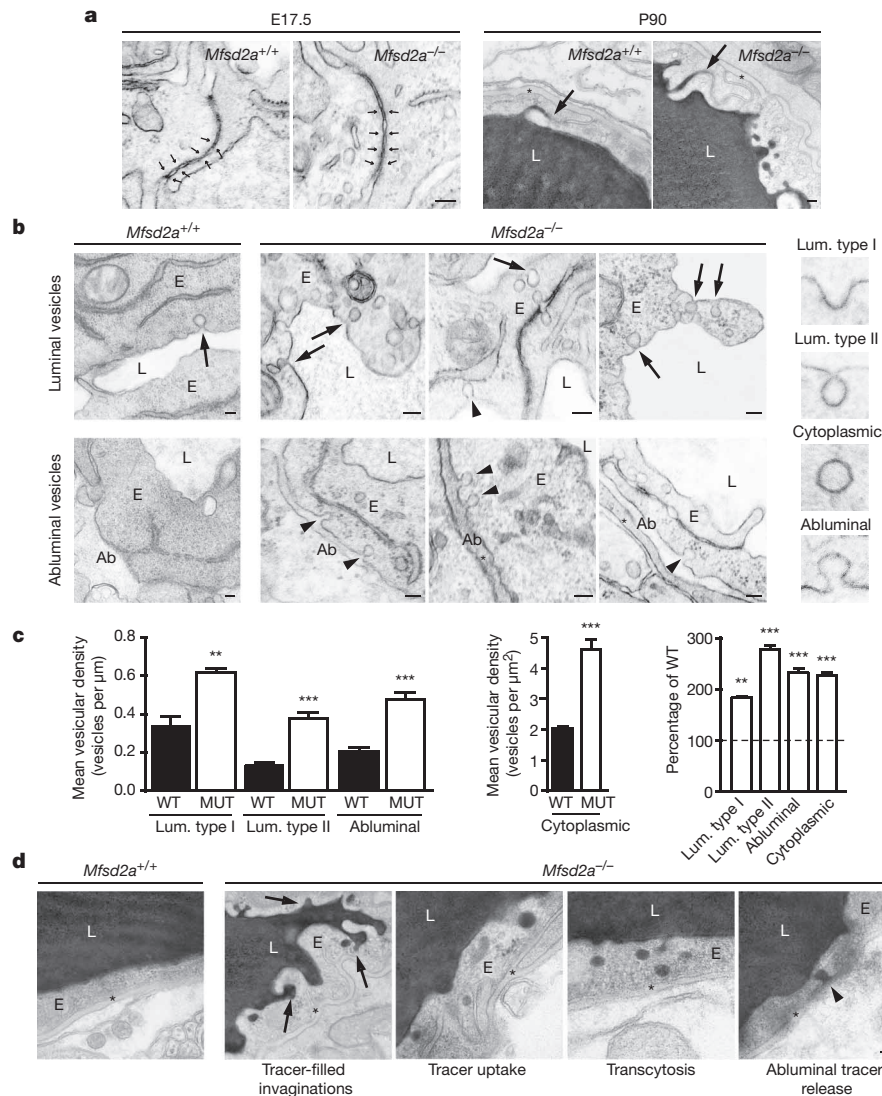


Figure 5 | *Mfsd2a* is required specifically to suppress transcytosis in brain endothelium to maintain BBB integrity. Electron-microscopy examination of BBB integrity. **a**, Embryonic *Mfsd2a*^{-/-} endothelium (E) showed no overt tight-junction ultrastructural defect (left, normal 'kissing points', small arrows). The vessel lumen (L) in HRP-injected adult mice was filled with electron-dense 3-3' diaminobenzidine (DAB) reaction (black) that diffused into intercellular clefts but stopped sharply at the junction without parenchymal leakage (right, arrows). **b**, Increased vesicular activity in embryonic *Mfsd2a*^{-/-} endothelium (E17.5). Left, wild-type endothelium displayed very few vesicles (arrow). Right, *Mfsd2a*^{-/-} endothelium contained many vesicles of various types:

luminal (arrows) and abluminal (Ab; arrowheads) membrane-connected and cytoplasmic vesicles. **c**, Vesicular density quantification (as shown in **b**, reference WT values (dashed line), see also Supplementary Fig. 7a). **d**, Increased transcytosis was evident in HRP-injected adult *Mfsd2a*^{-/-} mice (P90). In wild-type littermates (left) HRP activity was confined to the lumen with no HRP-filled vesicles. Many HRP-filled vesicles found in *Mfsd2a*^{-/-} endothelial cells (right, see quantification in Supplementary Fig. 7b). Luminal invaginations (dye uptake, arrows) and release to the basement membrane (abluminal side, asterisk). Scale bars, 100 nm (**a**, **b**), 200 nm (**c**). All data are mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$ (student's *t*-test).

profiles^{4,5}. Therefore we analysed published microarray data of two pericyte-deficient mouse models⁵ and found a dramatic downregulation of *Mfsd2a* in these mice, with a direct correlation between the reduction of *Mfsd2a* gene expression and the degree of pericyte coverage (Extended Data Fig. 7a). Furthermore, immunostaining for *Mfsd2a* in *Pdgfr*^{ret/ret} mice⁵ revealed a significant decrease in *Mfsd2a* protein levels in endothelial cells that are not covered by pericytes (Extended Data Fig. 7b–d). Therefore, it is plausible that the increased vesicular trafficking phenotype observed in pericyte-deficient mice is, at least in part, mediated by *Mfsd2a*, and that endothelial–pericyte interactions control the expression of *Mfsd2a*, which in turn controls BBB integrity.

We demonstrate that *Mfsd2a* is required to suppress endothelial transcytosis in the CNS. Because of *Mfsd2a*'s involvement in human trophoblast cell fusion²⁴ and of our genetic evidence for its role in suppressing transcytosis, we propose that *Mfsd2a* serves as a cell-surface molecule to regulate membrane fusion or trafficking. Indeed, from

immuno-electron-microscopy examination, *Mfsd2a* protein was found in the luminal plasma membrane and associated with vesicular structures in cerebral endothelial cells, but not in tight junctions (Extended Data Fig. 8). At present, it is not clear whether the reported transporter function of *Mfsd2a* is related to its role in BBB formation.

BBB breakdown has been reported in the aetiology of various neurological disorders^{7–10}, and two separate *Mfsd2a*-deficient mouse lines were reported to exhibit neurological abnormalities, such as ataxic behaviour^{21,25}. Finding a novel physiological role of *Mfsd2a* may provide a valuable tool to address how a non-functional BBB could affect brain development. In addition, our finding also highlights the importance of the transcytotic mechanism in BBB function, whereas most previous attention has been focused on potential BBB leaks through intercellular junctions. Indeed, increased numbers of pinocytotic vesicles were observed following acute exposure to external stress inducers in animal models²⁶, and have also been observed in human pathological conditions⁹. It will

be interesting to examine whether Mfsd2a is involved in these pathological and acute assault situations. We cannot be certain that the elevated levels of transcytosis in *Mfsd2a*^{-/-} mice were not due to some form of acute cellular stress, but it is very unlikely. This is because under stress, either cells respond to restore homeostasis or cell death occurs²⁷. However, increased transcytosis in *Mfsd2a*^{-/-} mice persists from embryonic stages to adulthood, and up to 6 months of age these mice exhibit no sign of vascular degeneration (Extended Data Fig. 5c). Our identification of a key molecular player in BBB formation may also aid efforts to develop therapeutic approaches for efficient drug delivery to the CNS. As an accessible cell surface molecule, Mfsd2a is poised to be a potential therapeutic target for BBB restoration and manipulation.

METHODS SUMMARY

The lowest volume of 10-kDa dextran tetramethylrhodamine, lysine fixable (D3312 Invitrogen) that still facilitated full perfusion was injected into the embryonic liver, while keeping the embryo connected to the maternal blood circulation through the umbilical cord. After 3 minutes of tracer circulation, embryonic heads were fixed by immersion in 4% paraformaldehyde (PFA) overnight at 4 °C, cryopreserved in 30% sucrose and frozen in TissueTek OCT (Sakura). Sections of 12 µm were then collected and post-fixed in 4% PFA at room temperature (20–25 °C) for 15 min, washed in PBS and co-stained with either α-PECAM antibody or with isolectin B4 to visualize blood vessels (see Methods for details). P90 HRP injection and E17.5 cortex capillaries transmission electron microscopy (TEM) imaging was done as described previously².

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2013; accepted 14 April 2014.

Published online 14 May 2014.

- Saunders, N. R., Liddel, S. A. & Dziegielewska, K. M. Barrier mechanisms in the developing brain. *Front. Pharmacol.* **3**, 46 (2012).
- Reese, T. S. & Karnovsky, M. J. Fine structural localization of a blood-brain barrier to exogenous peroxidase. *J. Cell Biol.* **34**, 207–217 (1967).
- Siegenthaler, J. A., Sohet, F. & Daneman, R. 'Sealing off the CNS': cellular and molecular regulation of blood-brain barrierogenesis. *Curr. Opin. Neurobiol.* **23**, 1057–1064 (2013).
- Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood-brain barrier integrity during embryogenesis. *Nature* **468**, 562–566 (2010).
- Armulik, A. *et al.* Pericytes regulate the blood-brain barrier. *Nature* **468**, 557–561 (2010).
- Bell, R. D. *et al.* Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 321–323 (2010).
- Zlokovic, B. V. The blood-brain barrier in health and chronic neurodegenerative disorders. *Neuron* **57**, 178–201 (2008).
- Zhong, Z. *et al.* ALS-causing SOD1 mutants generate vascular changes prior to motor neuron degeneration. *Nature Neurosci.* **11**, 420–422 (2008).
- Bell, R. D. & Zlokovic, B. V. Neurovascular mechanisms and blood-brain barrier disorder in Alzheimer's disease. *Acta Neuropathol.* **118**, 103–113 (2009).
- Bell, R. D. *et al.* Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512–516 (2012).
- Saunders, N. R. *et al.* Transporters of the blood-brain and blood-CSF interfaces in development and in the adult. *Mol. Aspects Med.* **34**, 742–752 (2013).
- Stenman, J. M. *et al.* Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science* **322**, 1247–1250 (2008).
- Liebner, S. *et al.* Wnt/β-catenin signaling controls development of the blood-brain barrier. *J. Cell Biol.* **183**, 409–417 (2008).
- Daneman, R. *et al.* Wnt/β-catenin signaling is required for CNS, but not non-CNS, angiogenesis. *Proc. Natl Acad. Sci. USA* **106**, 641–646 (2009).
- Tam, S. J. *et al.* Death receptors DR6 and TROY regulate brain vascular development. *Dev. Cell* **22**, 403–417 (2012).
- Cullen, M. *et al.* GPR124, an orphan G protein-coupled receptor, is required for CNS-specific vascularization and establishment of the blood-brain barrier. *Proc. Natl Acad. Sci. USA* **108**, 5759–5764 (2011).
- Wang, Y. *et al.* Norrin/Frizzled4 signaling in retinal vascular development and blood brain barrier plasticity. *Cell* **151**, 1332–1344 (2012).
- Alvarez, J. I. *et al.* The Hedgehog pathway promotes blood-brain barrier integrity and CNS immune quiescence. *Science* **334**, 1727–1731 (2011).
- Mizee, M. R. *et al.* Retinoic acid induces blood-brain barrier development. *J. Neurosci.* **33**, 1660–1671 (2013).
- Stern, L., Rapoport, J. L. & Lokschna, E. S. Le fonctionnement de la barrière hémato-encéphalique chez les nouveau-nés. *C. R. Soc. Biol.* **100**, 231–223 (1929).
- Tang, T. *et al.* A mouse knockout library for secreted and transmembrane proteins. *Nature Biotechnol.* **28**, 749–755 (2010).
- Esnault, C. placenta-specific receptor for the fusogenic, endogenous retrovirus-derived, human syncytin-2. *Proc. Natl Acad. Sci. USA* **105**, 17532–17537 (2008).
- Reiling, J. H. *et al.* A Haploid genetic screen identifies the major facilitator domain containing 2A (MFSD2A) transporter as a key mediator in the response to tunicamycin. *Proc. Natl Acad. Sci. USA* **108**, 11756–11765 (2011).
- Toufaily, C. *et al.* MFSD2a, the Syncytin-2 receptor, is important for trophoblast fusion. *Placenta* **34**, 85–88 (2013).
- Berger, J. H., Charron, M. J., Silver, D. L. Major facilitator superfamily domain-containing protein 2a (MFSD2A) has roles in body growth, motor function, and lipid metabolism. *PLoS One* **7**, e50629 (2012).
- Nag, S. In *The Blood-Brain Barrier—Biology and Research Protocols*. Part 2 (ed. Nag, S.) 99–100 (2003).
- Chen, F., Evans, A., Pham, J. & Plosky, B. (eds) *Mol. Cell Special Review issue* **40** (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Karnovsky, E. Raviola and T. Reese for advice and discussion; S. R. Datta, C. Weitz, M. Greenberg, Q. Ma, C. Harvey and members of the Gu laboratory for comments on the manuscript; D. Sabatini and J. Reeling for sharing unpublished data; C. Betsholtz and C. Olsson for providing *Pdgfr^{ret/ret}* mouse brain samples; T. Schwarz and A. Oztan for discussion and advice on cell trafficking; W.-J. Oh for help with graphic illustrations; the Flow Cytometry Facility in the department of Systems Biology at Harvard Medical School for cell sorting; the Microarray Core at Dana-Farber Cancer Institute for Affymetrix assay; HSPH Bioinformatics Core, Harvard School of Public Health, for assistance with microarray analysis and Gene Expression Omnibus (GEO) submission; the Enhanced Neuroimaging Core at Harvard NeuroDiscovery Center for helping with confocal imaging and image analysis; the HMS Electron Microscopy Core Facility, the Neurobiology Imaging Facility for consultation and instrument availability that supported this work (this facility is supported in part by the Neural Imaging Center as part of an NINDS P30 Core Center grant no. NS072030); R. Polakiewicz and J. Xie from CST for generating Mfsd2a antibodies. This work was supported by the Harold Perlman postdoctoral fellowships, the Goldenson postdoctoral fellowship, and the Lefler postdoctoral fellowship (A.B.-Z.); the DFG-German Research Foundation postdoctoral fellowship (E.K.); the Mahoney postdoctoral fellowship (B.L.); NIH training grant 5T32MH20017-15 (B.J.A.); and the Sloan research fellowship, Armenise junior faculty award, the Genise Goldenson fund, the Freudenberg award, and NIH grant R01NS064583 (C.G.).

Author Contributions C.G. and A.B.-Z. conceived and designed the project. A.B.-Z., B.L., E.K., B.J.A. and H.Y. performed experiments. A.B.-Z. and B.L. analysed data and performed image analysis and quantification. Y.M. analysed microarray data. C.G. and A.B.-Z. wrote the manuscript with significant input from B.L. and B.J.A.

Author Information Microarray data have been deposited in NCBI's Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO series accession number GSE56777. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.G. (chenghua_gu@hms.harvard.edu).

METHODS

Animals. Wild-type Swiss-Webster mice (Taconic Farms) were used for embryonic BBB functionality assays and expression profiles. Homozygous *Tie2-GFP* transgenic mice (Jackson laboratory, strain 003658) were used for BBB transcriptional profiling. *Mfsd2a*-null mice²¹ (Mouse Biology Program, University of California, Davis — MMRRC strain 032467-UCD, B6;129S5-Mfsd2atm1Lex/Mmucd) were maintained on C57Bl/6;129SVE mixed background and used for testing the involvement of *Mfsd2a* in barrier genesis. *Mfsd2a*-null mutant mice were genotyped using the following PCR primers: 5'-CCTGGTTGCTAAGTGCTAGC-3' and 5'-GTTCACTGGCTTGGAGGATGC-3', which provide a 210-bp product for the *Mfsd2a* wild-type allele; and 5'-CACTTCCTAAAGCCTTACTTC-3' and 5'-GCAGCGCATCGCCTTCTATC-3', which provide a 301-bp product for the *Mfsd2a*-knockout allele.

Pregnant mice were obtained following overnight mating (day of vaginal plug was defined as embryonic day 0.5).

All animals were treated according to institutional and US National Institutes of Health (NIH) guidelines approved by the Institutional Animal Care and Use Committee (IACUC) at Harvard Medical School.

Immunohistochemistry. Tissues were fixed with 4% paraformaldehyde (PFA) at 4 °C overnight, cryopreserved in 30% sucrose and frozen in TissueTek OCT (Sakura). Tissue sections were blocked with 5% goat serum, permeabilized with 0.5% Triton X-100, and stained with the following primary antibodies: α -PECAM (1:500; 553370, BD PharmingenTM), α -Claudin5 (1:400; 35-2500, Invitrogen), α -Mfsd2a (1:500; Cell Signaling Technologies (under development)), α -Pdgfr β (1:100; 141402, eBioscience), α -CD31 (1:100; 558744, BD PharmingenTM), α -SMA (1:100; C6198, Sigma Aldrich), followed by 568/488 Alexa Fluor-conjugated secondary antibodies (1:300–1:1000, Invitrogen) or with Isolectin B4 (1:500; I21411, Molecular Probes). Slides were mounted in Fluoromount G (EMS) and visualized by epifluorescence, light, or confocal microscopy.

In situ hybridization. Tissue samples were frozen in liquid nitrogen and embedded in TissueTek OCT (Sakura). Sections (18 μ m) were hybridized with a digoxigenin (DIG)-labelled mouse *Mfsd2a* antisense riboprobe (1,524–2,024 bp NM_029662) at 60 °C overnight. A sense probe was used to ensure signal specificity. For detection, signals were developed using anti-DIG antibody conjugated with alkaline phosphatase (Roche). After antibody treatment, sections were incubated with BM Purple AP Substrate (Roche).

Embryonic BBB permeability assay. The method is based on the well-established adult BBB dye-injection assay with special considerations for the injection site and volume to cater the nature of embryonic vasculature^{20,28–30}.

Four major modifications were made: first, embryos were injected while still attached via the umbilical cord to the mother's blood circulation, minimizing abrupt changes in blood flow. Deeply anaesthetized pregnant mice were used.

Second, taking advantage of the sinusoidal, fenestrated and most permeable liver vasculature, dye was injected using a Hamilton syringe into the embryonic liver and was taken into the circulation in a matter of seconds.

Third, dye volume was adjusted to a minimum that still allows detection in all CNS capillaries after 3 min of circulation. High-fluorescence intensity dye enables the use of small volumes and facilitates detection at the single-capillary level (10-kDa dextran-tetramethylrhodamine, lysine fixable, 4 mg ml⁻¹ (D3312 Invitrogen), 1 μ l for E13.5, 2 μ l for E14.5, 5 μ l for E15.5).

Fourth, traditional perfusion fixation was omitted, again to prevent damage to capillaries. Instead, fixable dyes were used to allow reliable immobilization of the dye at the end of the circulation time (relatively small embryonic brain facilitates immersion fixation).

Embryonic heads were fixed by immersion in 4% PFA overnight at 4 °C, cryopreserved in 30% sucrose and frozen in TissueTek OCT (Sakura). Sections of 12 μ m were then collected and post-fixed in 4% PFA at room temperature for 15 min, washed in PBS and co-stained with either α -PECAM antibody or with isolectin B4 to visualize blood vessels. All embryos from each litter were injected blind before genotyping.

Postnatal and adult BBB permeability assay. P2–P5 pups were deeply anaesthetized and three methods were used: the first method involved injection of 10 μ l of 10-kDa or 70-kDa dextran tetramethylrhodamine (4 mg ml⁻¹ D3312 Invitrogen) into the left ventricle with a Hamilton syringe. After 5 min of circulation, brains were dissected and fixed by immersion in 4% PFA at 4 °C overnight, cryopreserved in 30% sucrose and frozen in TissueTek OCT (Sakura). Sections of 12 μ m were collected and post-fixed in 4% PFA at room temperature for 15 min, washed in PBS and co-stained to visualize blood vessels with either α -PECAM primary antibody (1:500; 553370, BD Pharmingen), followed by 488-Alexa Fluor conjugated secondary antibody (1:1000, Invitrogen) or with isolectin B4 (1:500; I21411, Molecular Probes).

The second method involved injection of 10 μ l of HRP type II (5 mg ml⁻¹ P8250-50KU Sigma-Aldrich) into the left heart ventricle with a Hamilton syringe.

After 5 min of circulation brains were dissected and immersion fixed in 2% glutaraldehyde in 4% PFA in cacodylate buffer (0.1 M, pH 7.3) at room temperature for 1 h then at 4 °C for 3 h then washed in cacodylate buffer overnight. Cortical-vibratome sections (100 μ m) were processed in a standard DAB reaction.

The third method involved the use of EZ-link NHS-sulfo-biotin as a tracer, as described previously¹⁷.

Imaging. Nikon Eclipse 80i microscope equipped with a Nikon DS-2 digital camera was used to image HRP tracer experiments, vasculature density and pericyte coverage comparisons and expression analyses. Zeiss LSM 510 META upright confocal microscope was used to image Dextran and NHS-sulfo-biotin BBB permeability assays. A Nikon FluoView FV1000 laser scanning confocal microscope and a Leica SP8 laser scanning confocal microscope were used for imaging *Mfsd2a* and pericyte marker immunohistochemistry. Images were processed using Adobe Photoshop and ImageJ (NIH).

Morphometric analysis of vasculature. Coronal sections (25- μ m thick) of E15.5, P4 and P70 brains were immunostained for PECAM. For vascular density and branching, confocal images were acquired with a Nikon FluoView FV1000 laser scanning confocal microscope and maximal projection images (5 per animal) were used for quantifications. The number of branching points was manually counted. Capillary density was quantified using MetaMorph software (Universal Imaging, Downingtown, Pennsylvania) by measuring the area occupied by PECAM-positive vessels per cortical area. The mean capillary diameter was measured manually in ImageJ from cross-sectional vascular profiles (20 per animal) on micrographs (5–7 per animal) taken under a $\times 60$ objective with a $\times 2$ digital zoom.

For artery distribution quantification, 25- μ m-thick sections (P60) were stained for smooth muscle actin (SMA) and PECAM. The proportion of PECAM-positive brain vessels with artery (SMA) identity was quantified using MetaMorph and expressed as percent of controls. Quantification was carried out blind.

Quantification of cortical-vessel pericyte coverage. Pericyte coverage of cortex vessels in *Mfsd2a*^{-/-} and wild-type littermate control mice was quantified by analysing the proportion of total claudin-5-positive endothelial length also positive for the pericyte markers CD13 or Pdgfr β . Immunostaining was performed on 20- μ m sections of P5 cortex. In each animal, 20 images of 10 different sections were analysed. Microvasculature was found to be completely covered by pericytes in both control and *Mfsd2a*^{-/-} mice and therefore no error bars are presented for the average pericyte coverage in Extended Data Fig. 6a, b ($n = 3$). All the analysis was done with ImageJ (NIH). Quantification was carried out blind.

Quantification of vessel leakage. Epifluorescence images of sections from injected tracer and co-stained with lectin were analysed manually with ImageJ (NIH). Coronal cortical sections (12 μ m) of the same rostrocaudal position were used for the analysis. The same acquisition parameters were applied to all images and the same threshold was used. Tracer-positive cells found outside a vessel (parenchyma) were used as a parameter for leakage. For each embryo, at least 20 sections of a fixed lateral cortical plate area were scored. Four arbitrary leakage groups were classified based on the number of tracer parenchyma positive cells per section (0, 1–5, 5–10 and 10–40). Average representation of each leakage group was calculated for *Mfsd2a*^{-/-} and control embryos. Quantification was carried out blind.

Spectrophotometric quantification of 10-kDa fluoro-ruby-dextran tracer was carried out from cortical extracts, 16 h after tail-vein injections in adult mice, as described previously⁵.

Transmission electron microscopy. TEM imaging of P90 HRP injection and E17.5 cortex capillaries was carried out as described previously². HRP (10 mg (per 20 g); Sigma Aldrich, HRP type II) were dissolved in 0.4 ml of PBS and injected into the tail veins of deeply anaesthetized P90 mice. After 30 min of HRP circulation, brains were dissected and fixed by immersion in a 0.1 M sodium-cacodylate-buffered mixture (5% glutaraldehyde and 4% PFA) for 1 h at room temperature followed by 5 h in PFA at 4 °C. Following fixation, the tissue was washed overnight in 0.1 M sodium-cacodylate buffer and then cut in 50- μ m-thick free-floating sections using a vibratome. Sections were incubated for 45 min at room temperature in 0.05 M Tris-HCl pH 7.6 buffer, containing 5.0 mg per 10 ml of 3-3' diaminobenzidine (DAB, Sigma Aldrich) with 0.01% hydrogen peroxide. Sections were then post-fixed in 1% osmium tetroxide and 1.5% potassium ferrocyanide and dehydrated and embedded in epoxy resin. E17.5 samples were processed as the P90 samples without HRP injection and with longer fixation times (2–3 days in room temperature). Ultrathin sections (80 nm) were then cut from the block surface, collected on copper grids, stained with Reynold's lead citrate and examined under a 1200EX electron microscope (JEOL) equipped with a 2k CCD digital camera (AMT).

Immunogold labelling for electron microscopy. Mice were deeply anaesthetized and perfused through the heart with 30 ml of PBS followed by 150 ml of a fixative solution (0.5% glutaraldehyde in 4% PFA prepared in 0.1 mM phosphate buffer, pH 7.4), and then by 100 ml of 4% PFA in phosphate buffer. The brain was removed and post fixed in 4% PFA (30 min, 4 °C) and washed in PBS. Coronal brain sections (50- μ m thick) were cut on the same day with a vibratome and processed free floating.

Sections were immersed in 0.1% sodium borohydride in PBS (20 min, room temperature), rinsed in PBS and pre-incubated (2 h) in a blocking solution of PBS containing 10% normal goat serum, 0.5% gelatine and 0.01% Triton. Incubation (24 h, 20–25 °C) with rabbit anti-Mfsd2a (1:100; Cell Signaling Technologies (under development)) primary antibody was followed by rinses in PBS and incubation (overnight, 20–25 °C) in a dilution of gold-labelled goat anti-rabbit IgGs (1:50; 2004, Nanoprobes). After washes in PBS and sodium acetate, the size of immunogold particles was silver-enhanced and sections rinsed in phosphate buffer before processing for electron microscopy.

Statistical analysis. Comparison between wild-type and *Mfsd2a*^{−/−} pericyte coverage and spectrophotometric quantification of 10-kDa fluoro-ruby-dextran tracer leakage was performed by a Mann–Whitney *U*-test (appropriate for small sample size; each embryo was considered as a sample). An unpaired student's *t*-test was used (GraphPad Prism 4 Software) for comparison between wild-type and *Mfsd2a*^{−/−} for vascular density, artery distribution, number of vesicular types, mean capillary diameter and Mfsd2a expression in pericyte deficient mice. *P* < 0.05 was considered significant (StatXact Cytel Software Corporation, Cambridge, Massachusetts, USA).

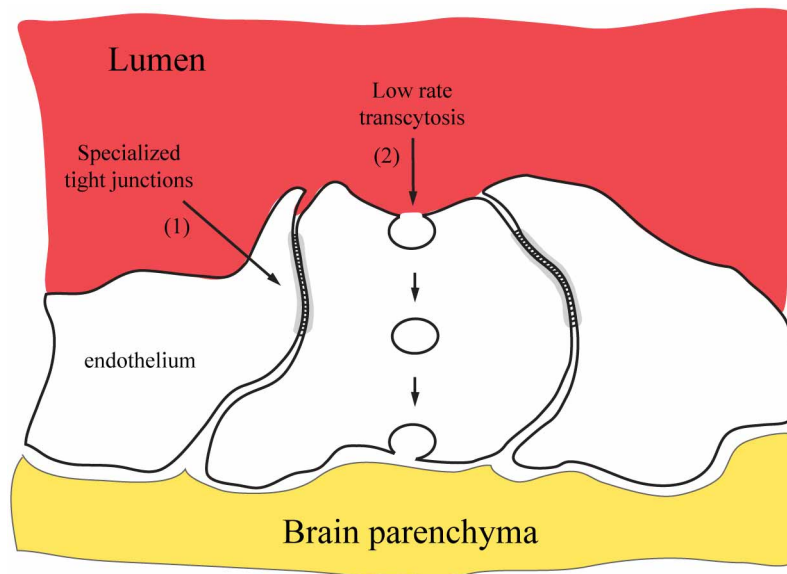
Transcriptional profiling. E13.5 *Tie2-GFP* embryos were micro-dissected for cortex and lungs. Cortex tissue was carefully cleared of the meninges and choroid plexus. FACS purification of GFP-positive cells and GeneChip analysis was performed as described previously³¹. RNA was purified with Arcturus PicoPure RNA isolation kit (Applied biosystems), followed by NuGEN Ovation V2 standard linear amplification and hybridization to Affymetrix Mouse Genome 430 2.0 Array. All material from a single litter (10–13 embryos) was pooled and considered as a biological replicate. Four biological replicates were used. Each biological replicate represents purification from different litters performed on different days.

Transcriptional profile analysis of pericyte deficient mice. Expression data from a published study of pericyte-deficient mice⁵ were obtained from the Gene Expression

Omnibus (<http://www.ncbi.nlm.nih.gov/geo>, accession number GSE15892). All microarrays were analysed using the MAS5 probe set condensation algorithm with Expression Console software (Affymetrix). *P* values were determined using a two-tailed student's *t*-test (*n* = 4).

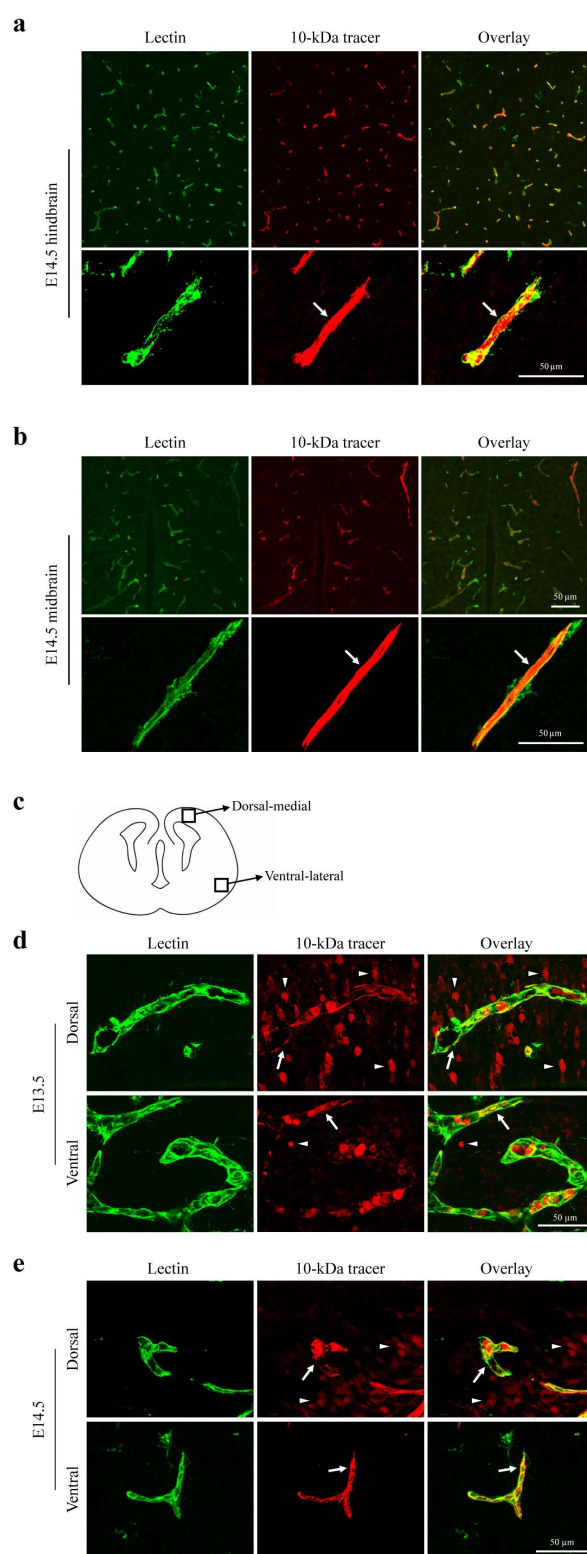
Mfsd2a protein expression in *Pdgfr*^{ret/ret} mice. Brain samples from P10–P14 mice and controls were kindly provided by C. Betsholtz. Sample processing and immunohistochemistry was carried out as described for all other samples in our study. Mfsd2a staining quantification was carried out with 12-μm cortical sagittal sections. Confocal images were acquired with a Nikon FluoView FV1000 laser scanning confocal microscope. Quantification of mean grey value per vascular profile was done with ImageJ (NIH) by outlining vascular profiles according to lectin staining and measuring Mfsd2a intensity in these areas. In all images, *Pdgfr*β antibody staining was used to test presence of pericytes in quantified vessels. *n* = 2 animals per genotype, 60 images quantified of at least 600 vascular profiles per animal. Quantification was carried out blind.

28. Ek, C. J., Habgood, M. D., Dziegielewska, K. M. & Saunders, N. R. Functional effectiveness of the blood-brain barrier to small water-soluble molecules in developing and adult opossum (*Monodelphis domestica*). *J. Comp. Neurol.* **496**, 13–26 (2006).
29. Risau, W., Hallmann, R. & Albrecht, U. Differentiation-dependent expression of proteins in brain endothelium during development of the blood-brain barrier. *Dev. Biol.* **117**, 537–545 (1986).
30. Bauer, H. *et al.* Ontogenic expression of the erythroid-type glucose transporter (Glut 1) in the telencephalon of the mouse: correlation to the tightening of the blood-brain barrier. *Brain Res. Dev. Brain Res.* **86**, 317–325 (1995).
31. Daneman, R. *et al.* The mouse blood-brain barrier transcriptome: a new resource for understanding the development and function of brain endothelial cells. *PLoS One* **5**, 313741 (2010).

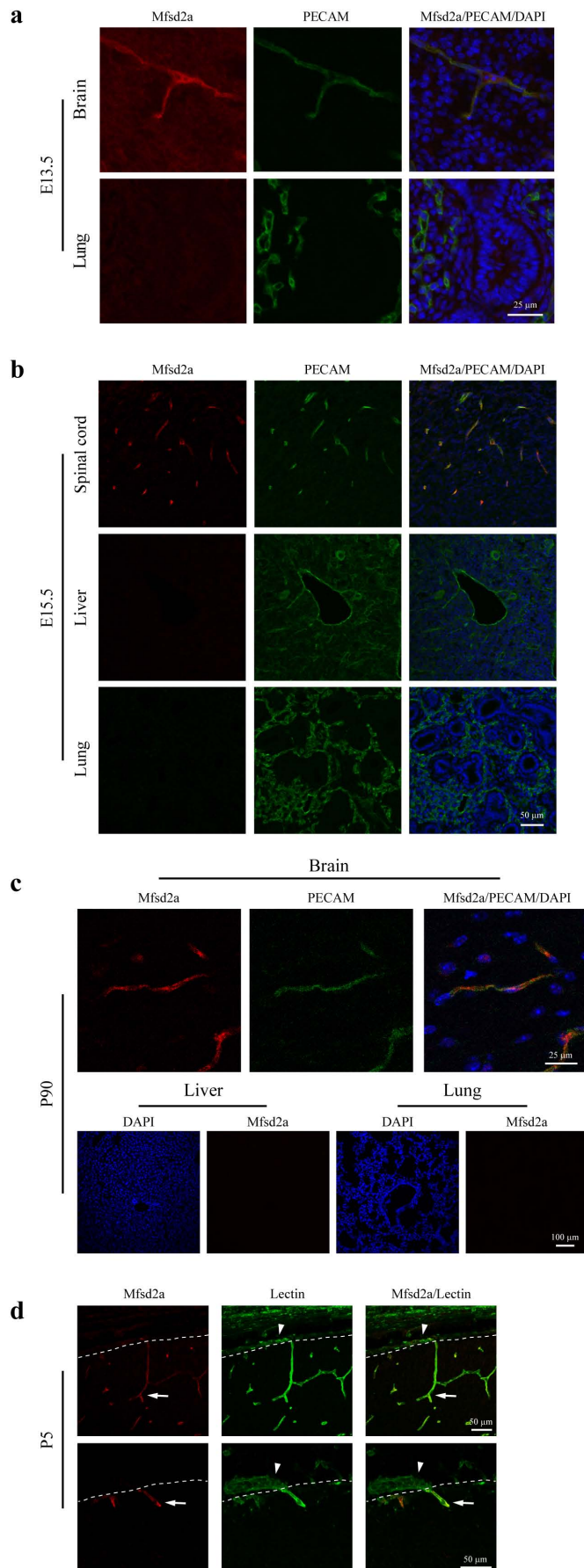


Extended Data Figure 1 | Diagram illustrating two unique BBB properties of CNS endothelial cells. Compared to the endothelial cells from the rest of the body, CNS endothelial cells that possess a BBB are characterized by highly

specialized tight junctions sealing the space between adjacent cells (1), and an unusually low rate of transcytosis from the vessel lumen to the brain parenchyma (2).

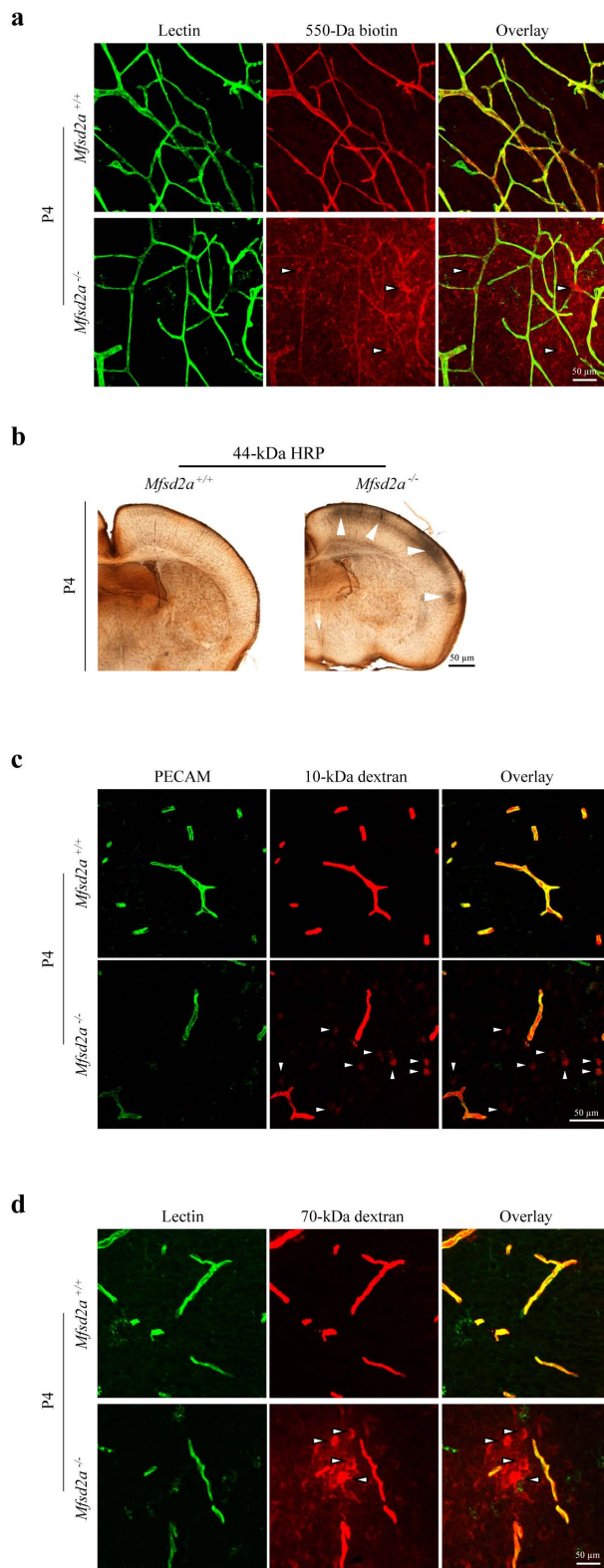


Extended Data Figure 2 | Spatial and temporal BBB maturation across brain regions and cortical regions. **a, b,** Embryonic BBB develops in a caudal-to-rostral spatial pattern. Dextran-tracer (10 kDa) injection revealed that the BBB is already functional in the hindbrain (**a**) and midbrain (**b**) at E14.5, a time-point at which cortical BBB is still leaky. Epifluorescence (low magnification) and confocal (high magnification) images of brain sections from injected embryos are shown. As illustrated in both **a** and **b**, tracer was confined to blood vessels (arrows). **c–e,** Cortical BBB develops in a ventrolateral to dorsomedial spatial pattern. **c,** Diagram of the embryonic cortex indicating dorsal–medial and ventral–lateral cortical regions illustrated in **d** and **e**. **d,** At E13.5, the BBB of the dorsal cortex (top panels) is not fully formed, as evidenced by little tracer inside the blood vessels (arrow) and tracer-filled parenchymal cells (arrowheads). In the ventral cortex (bottom panels), capillaries were better sealed, showing more tracer within the lumen (arrow) and less tracer in the brain parenchyma (arrowhead). **e,** At E14.5, the capillaries in the dorsal cortex (top panels) were still leaky, with little tracer inside the capillaries (arrow) and tracer-stained surrounding parenchyma (arrowheads). At the same age, the BBB of the ventral cortex (bottom panels) was already functional, with all tracer confined to the capillaries (arrow). Green, lectin; red, 10-kDa tracer. $n = 6$ embryos from 3 litters for each age.



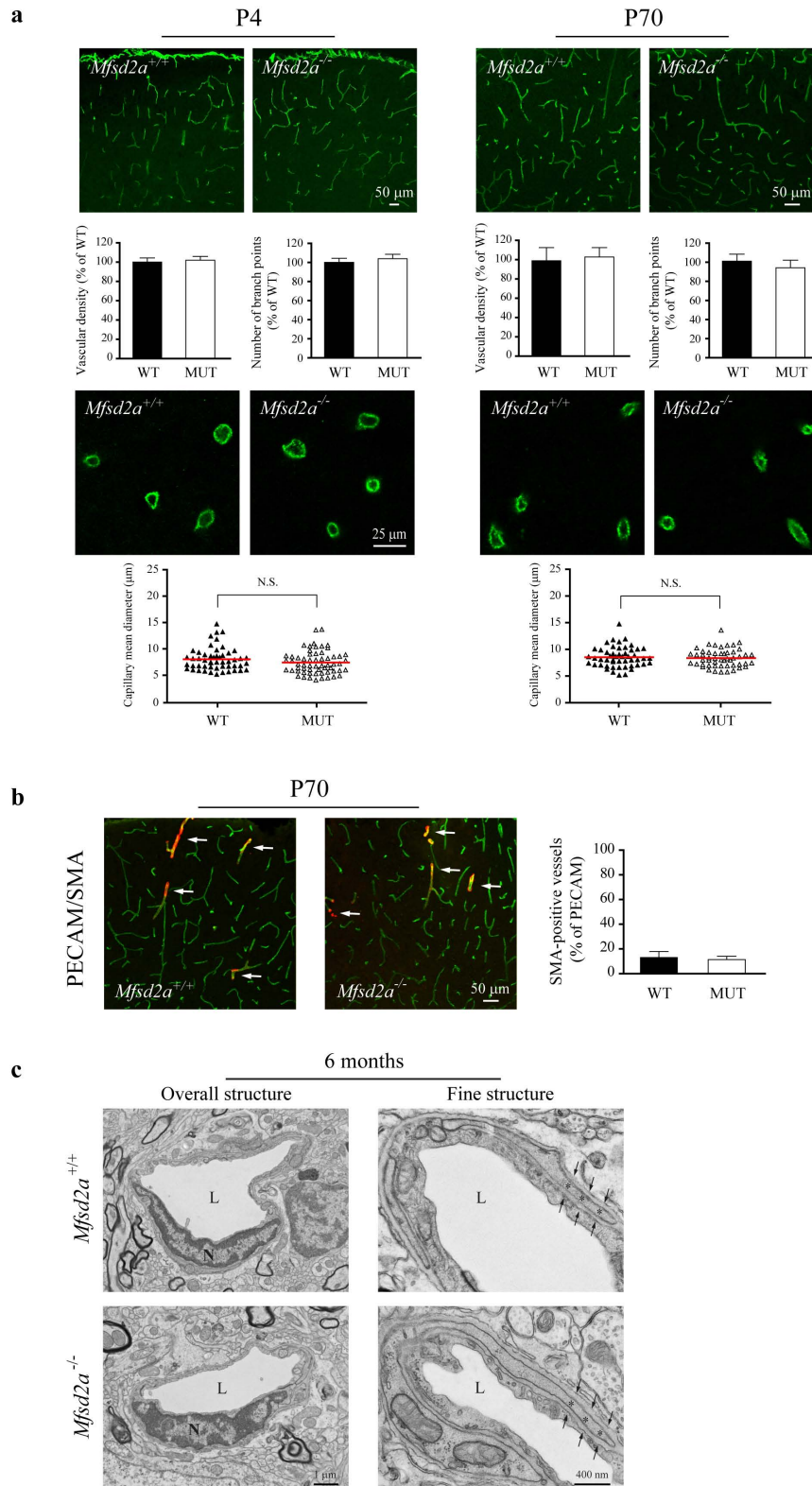
Extended Data Figure 3 | Mfsd2a protein is selectively expressed in BBB-containing CNS vasculature of both embryos and adults.

Immunohistochemical staining of Mfsd2a protein demonstrating its specific expression in BBB-containing CNS vasculature. Red, Mfsd2a; green, PECAM (endothelium); blue, DAPI (nuclei). **a**, At E13.5, the time of BBB establishment, Mfsd2a expression was detected in the brain (top panels) but not in the lung vessels (bottom panels), confirming the microarray data. **b**, At E15.5, the first developmental time-point of BBB functionality, Mfsd2a expression was detected in spinal cord (top panels) but not in lung (middle panels) or liver vessels (bottom panels). **c**, Mfsd2a selective expression in BBB-containing vessels persists in adult mice (P90) as shown in brain vessels (top panels) but not in lung or liver (bottom panels). **d**, Mfsd2a is expressed in cerebral vessel (arrow) but not in pial vessels (arrow heads). Low (top panel) and high magnification (bottom panel) of pial-cerebral boundary (dotted line) of P5 dorsal cortex (wild-type mice). $n = 3$ embryos from 3 litters for each age.



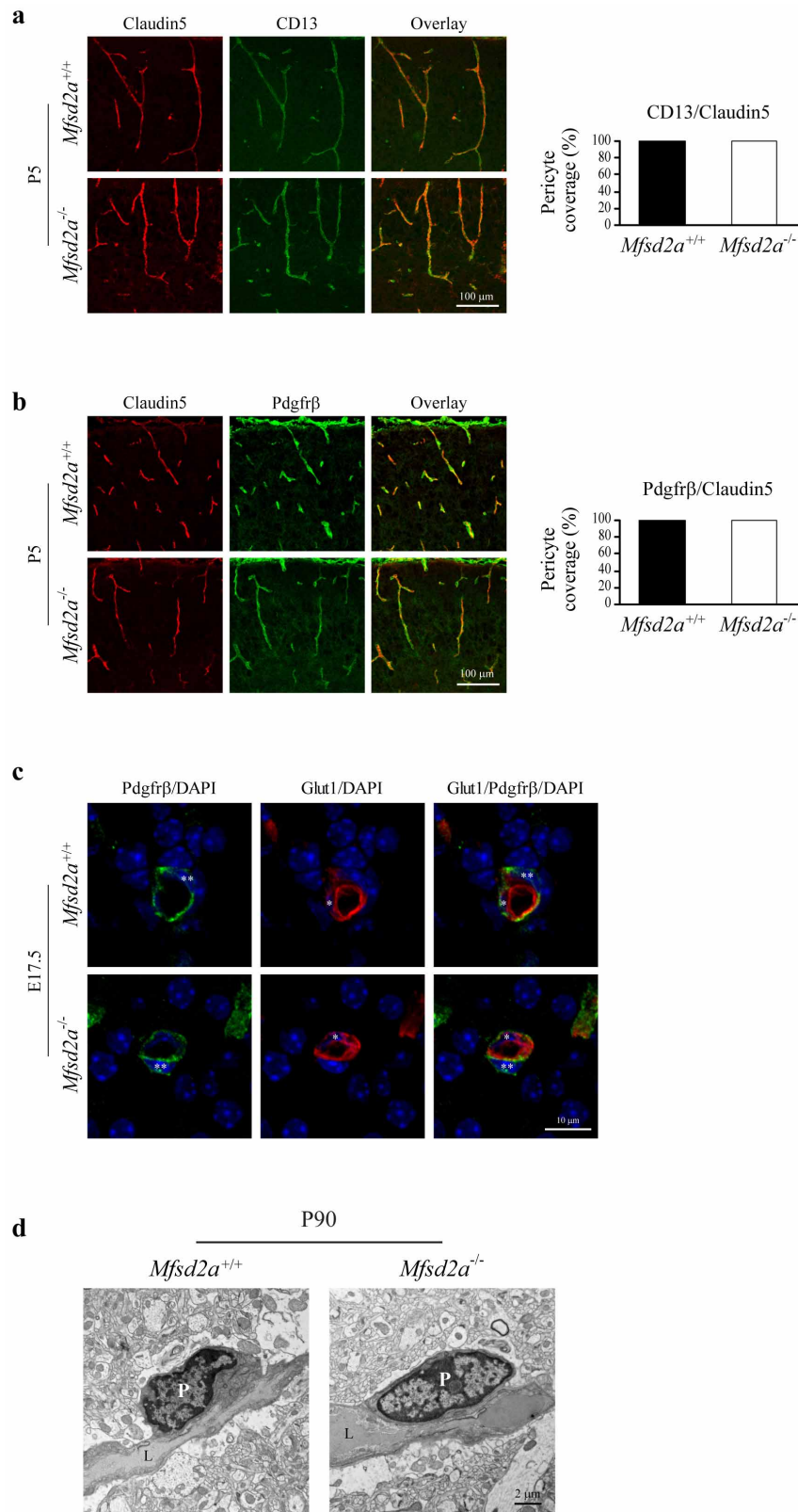
Extended Data Figure 4 | The leaky BBB phenotype in *Mfsd2a*^{-/-} mice persists after birth and is not restricted to carbohydrate-based tracers.

a, b, Injection of two non-carbohydrate-based tracers with different molecular weight and different molecular compositions at P4 revealed a persistent leaky barrier phenotype in mice lacking *Mfsd2a*. **a**, The small-molecular-weight tracer sulfo-NHS-biotin (~550 Da), was confined to vessels in wild-type controls (upper panels), whereas it leaked out of the vessels (arrowheads) in *Mfsd2a*^{-/-} mice (lower panels). Green, lectin; red, tracer. **b**, The high-molecular-weight protein tracer HRP (~44 kDa) was confined to vessels in control mice at P4 (left), whereas in mice lacking *Mfsd2a* (right), tracer was diffusing into the brain parenchyma (arrowheads). Brown, visualization of HRP in light microscopy by DAB reaction. Stainings were carried out on 100-μm cortical sections of tracer-injected animals. *n* = 4 pups per genotype from 3 litters. **c, d**, BBB leakage of tracer with 70 kDa molecular weight is observed in postnatal *Mfsd2a*^{-/-} mice. Green, lectin or PECAM (vessels); red, tracer. The 10-kDa (**a**) and 70-kDa (**b**) tracers fluoro-ruby-dextran were confined to vessels in control mice (top panels), but not in *Mfsd2a*^{-/-} mice (bottom panels), where tracer was taken up by brain parenchymal cells (arrowheads). Stainings were carried out on 12-μm cortical sections of tracer-injected animals. *n* = 3 pups per genotype from 3 litters.



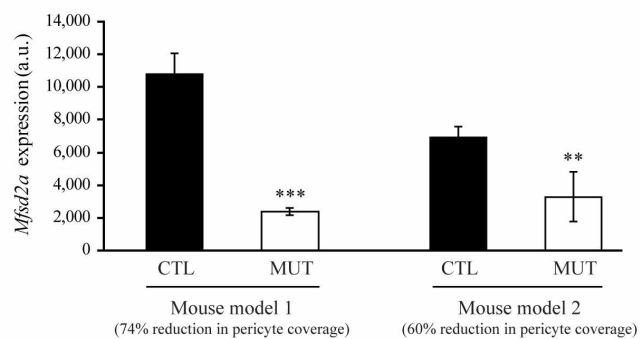
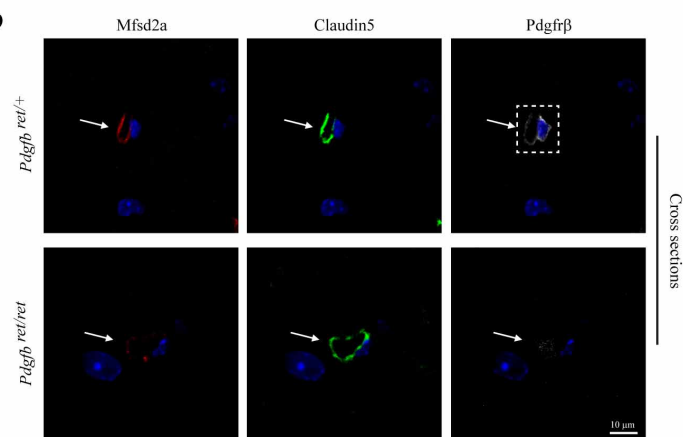
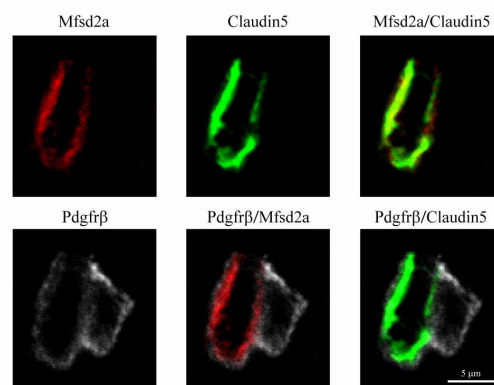
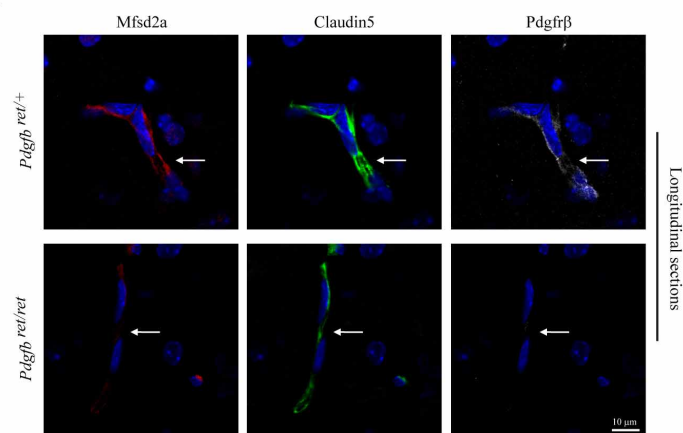
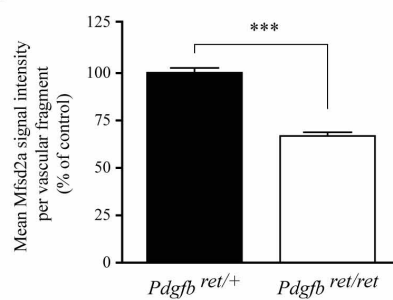
Extended Data Figure 5 | Perinatal and adult mice lacking *Mfsd2a* do not display changes in cerebrovascular network properties or signs of vascular degeneration. **a**, No abnormalities were found in cortical capillary density and vascular branching (top panels) as well as capillary diameter (bottom panels) of postnatal (P4, left) and adult (P70, right) *Mfsd2a*^{-/-} mice. Images of vascular staining in coronal cortical sections and high-magnification images of cross-section profiles of capillaries (green, PECAM) and quantification. Data are mean \pm s.e.m. $n = 3$ animals per genotype, 20 sections per animal. **b**, Immunostaining for smooth-muscle actin (arterial identifier, arrows) revealed no abnormalities in arterial distribution and specification in

Mfsd2a^{-/-} mice. Images of coronal cortical sections (left, green, PECAM; red, SMA) and quantification (right). Data are mean \pm s.e.m. $n = 3$ animals per genotype, 20 sections per animal. **c**, Electron-microscopy examination of older *Mfsd2a*^{-/-} mice did not reveal signs of cerebrovascular degeneration. Left, the overall capillary structure (for example, cell size, shape of the nucleus, thickness of the vessel wall, basement membrane integrity and pericyte attachment) did not differ between wild-type and mutant mice. Right, at higher magnification, normal features, such as pericyte (asterisk) attachment within a normal basement membrane (between arrows), could be observed in mice lacking *Mfsd2a*.



Extended Data Figure 6 | Pericyte coverage, attachment and ultrastructure are normal in *Mfsd2a*^{-/-} mice. **a, b,** *Mfsd2a*^{-/-} mice exhibit normal pericyte coverage. Co-staining of endothelium (claudin-5, red) and pericytes (CD13 in **a** and Pdgfrβ in **b**, green) revealed no overt difference in pericyte coverage of dorsal cortex vessels between wild-type (top row) and *Mfsd2a*^{-/-} (bottom row) mice at P5. Quantification of vascular coverage in both **a** and **b** showed no significant difference between wild-type and *Mfsd2a*^{-/-} samples ($P > 0.5$). Data are mean \pm s.e.m. $n = 3$ pups per genotype, 20 sections per animal. **c,** E17.5 *Mfsd2a*^{-/-} mice exhibit normal pericyte-endothelial attachment.

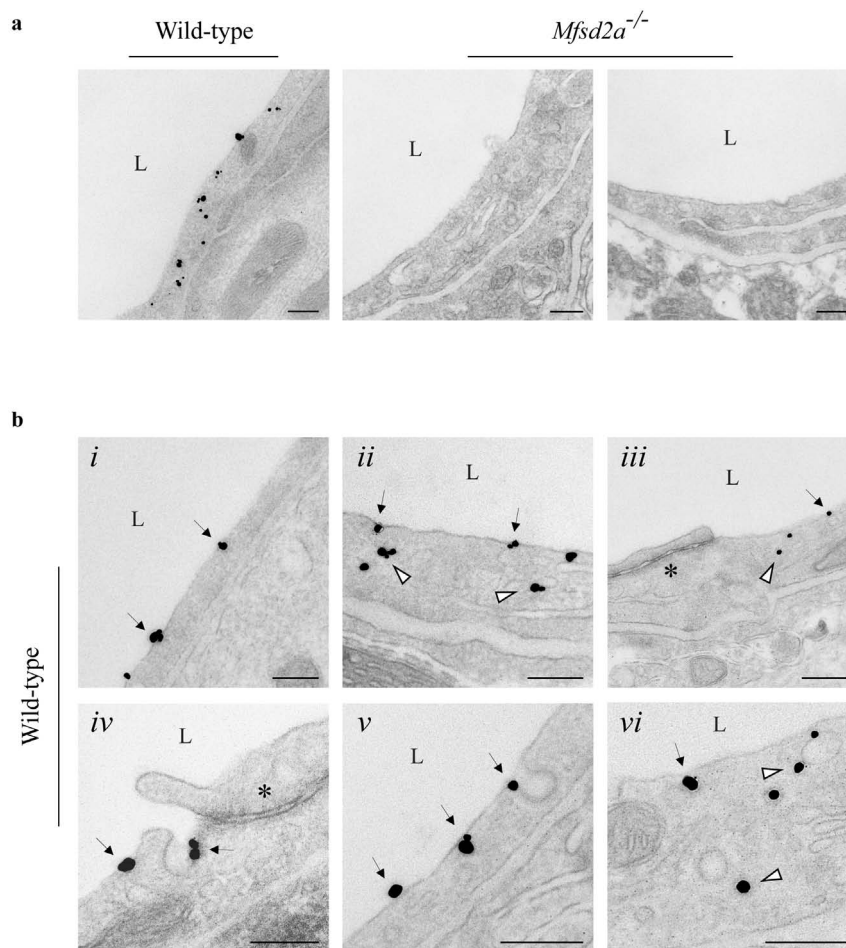
High magnification images of capillary cross sections co-staining for endothelium (Glut1, red) and pericytes (Pdgfrβ, green) revealed no difference in pericyte-endothelial relationships between wild-type (top panels) and *Mfsd2a*^{-/-} (bottom panels) mice (endothelial nucleus and pericyte nucleus are indicated by single and double asterisks, respectively). **d,** Electron micrographs of longitudinal capillary sections revealed that pericytes had normal appearance and were well positioned on the vessel walls in *Mfsd2a*^{-/-} adult mice; pericytes were adjacent to endothelial cells and shared a common basement membrane. L, lumen; P, pericyte.

a**b****b'****c****d**

Extended Data Figure 7 | *Mfsd2a* gene expression and Mfsd2a protein levels are downregulated in mouse models of reduced pericyte coverage.

a, Analysis of microarray data⁵ showed high levels of *Mfsd2a* expression in wild-type adult brain microvasculature, but a significant decrease in levels of *Mfsd2a* expression in mice that have reduced pericyte coverage at the BBB. *Pdgfr^{ret/ret}* mice (mouse model 1), where Pdgfr β binding to heparan sulphate proteoglycans was disrupted, exhibited a major loss of pericyte coverage (74% reduction)⁵ and showed a dramatic decrease in *Mfsd2a* expression (74% reduction) compared to that of littermate controls. Similarly, *Tie2^{Cre}/R26^{P+/0}*, *Pdgfr^{-/-}* mice (mouse model 2) in which *Pdgfr*-null alleles were complemented by one copy of human *PDGFB* transgene showed a less dramatic loss of pericyte coverage (60% reduction)⁵ and a smaller decrease in *Mfsd2a* expression (53% reduction). ** $P = 0.004$, *** $P = 1 \times 10^{-5}$). Bars reflect normalized signal of the *Mfsd2a* probe (1428223_at) in adult brain or cortex microvascular fragments (a.u.). Data are mean \pm s.d. of 4 biological replicates. **b**, Mfsd2a immunostaining in cerebrovasculature of postnatal

Pdgfr^{ret/ret} mice and littermate controls (*Pdgfr^{ret/+}*) revealed reduced Mfsd2a protein expression in endothelial cells that are not covered by pericytes. Cross section: red, Mfsd2a; green, claudin-5 (endothelium); blue, DAPI (nuclei); grey, Pdgfr β (pericyte). Reduced Mfsd2a signal was observed in endothelial cells of capillaries not covered with pericytes in *Pdgfr^{ret/ret}* mice (arrow, bottom panels), whereas strong Mfsd2a signal was apparent in pericyte-covered capillaries in *Pdgfr^{ret/+}* mice (arrow, top panel). Insets in **b'** demonstrate in high magnification that Mfsd2a expression is restricted to endothelial cells (co-localization with claudin-5 staining) and absent in pericytes (none co-localization with Pdgfr β staining). **c**, Additional example of the reduction in Mfsd2a expression in endothelial cells of *Pdgfr^{ret/ret}* mice (longitudinal section). **d**, Quantification of mean fluorescence intensity per vascular profile showed significant reduction of Mfsd2a signal in *Pdgfr^{ret/ret}* capillaries compared to controls. Data are mean \pm s.e.m. $n = 2$ animals per genotype, 60 images quantified of at least 600 vascular profiles per animal.



Extended Data Figure 8 | Immuno-electron-microscopy reveals the subcellular localization of Mfsd2a on the plasma membrane and vesicles, but not in tight junctions of cerebral endothelial cells. **a**, Electron micrographs showing silver-enhanced immunogold labelling of Mfsd2a in cerebral cortex capillaries from wild-type (left) but not in *Mfsd2a*^{-/-} mice (right), demonstrating staining specificity. **b**, Top panels (*i-iii*), three representative examples of Mfsd2a localization on the plasma membrane

(arrows) and in the cytoplasm (arrowheads), but not in tight junctions (asterisk). Bottom panels (*iv-vi*), high magnification representative examples of Mfsd2a localization on the luminal plasma membrane (arrows), associated with luminal invaginating vesicles (*iv, v*) and with cytoplasmic vesicles (arrowheads). All samples are of cortical vessels from adult mice (P30-P90). *n* = 2 for each genotype. L, lumen.

Extended Data Table 1 | Expression-profile comparison of cortex (BBB) and lung (non-BBB) endothelial cells

a

Probe set	Gene Symbol	Cortex GFP+	Lung GFP+	Cortex/Lung
1418788_at	Tie2	2765.0	5040.7	0.5
1417839_at	Cldn5	4978.0	4709.7	1.1
1440926_at	flt1	771.5	606.1	1.3
1449379_at	Vegfr2	3578.6	2115.3	1.7
1433956_at	VE-cadherin	2422.9	3572.4	0.7
1421287_a_at	PECAM	364.4	725.4	0.5

b

Probe set	Gene Symbol	Cortex GFP+	Lung GFP+	Cortex/Lung
1423343_at	Slco1c1	2918.1	51.7	56.4
1454622_at	Slc38a5	3007.5	124.7	24.1
1435418_at	Slc22a8	253.6	11.28	22.5
1448299_at	Slc1a1	565.9	47.4	11.9
1418706_at	Slc38a3	917.2	117.9	7.8
1418326_at	Slc7a5	5913.1	1050.9	5.6
1433933_s_at	Slco2b1	994.3	207	4.8
1447227_at	Slc40a1	81.8	20.3	4.0
1436137_at	Slc6a17	567	147.5	3.8
1436417_at	Slc19a3	124.6	34.4	3.6
1451486_at	Slc46a3	371.3	112.2	3.3
1422906_at	Abcg2	458.6	166.2	2.8
1436164_at	Slc30a1	2768	1178.4	2.3
1434773_a_at	Slc2a1	9337.6	4144.4	2.3
1419759_at	Abcb1a	1520	676.8	2.2
1454991_at	Slc7a1	8403.4	4403.5	1.9
1429154_at	Slc35f2	369.6	211	1.8
1433751_at	Slc39a10	3222.8	2074.3	1.6
1437149_at	Slc6a6	2096.5	1367.6	1.5
1433750_at	Slc31a1	1666.4	1389.2	1.2
1447851_x_at	Atp10a	328.4	276.6	1.2
1443870_at	Abcc4	138.4	117.5	1.2
1424211_at	Slc25a33	99.2	93	1.1
1426082_a_at	Slc16a4	94.3	95.8	1.0
1418445_at	Slc16a2	1953.4	2069.4	0.9
1415802_at	Slc16a1	925.1	989.5	0.9
1423109_s_at	Slc25a20	310.4	410.7	0.8
1417022_at	Slc7a3	51.5	74.4	0.7
1453149_at	Slc25a32	109.2	204.7	0.5

a, Pan-endothelial markers were highly expressed in endothelial cells isolated from both cortex and lung of the endothelial specific *Tie2-GFP* reporter mouse at E13.5. **b**, Genes involved in the transport of molecules across the BBB, known as adult BBB markers, were highly enriched in brain versus lung endothelial cells. Data are mean of 4 biological replicates (4 litters).

Extended Data Table 2 | Vesicular activity and transcytosis events in the brain endothelial cells of *Mfsd2a*^{-/-} mice are dramatically increased**a**

Density of vesicles in the embryo brain endothelium (E17.5)

Tissue source	No. of endothelial profiles	No. of vesicles	Mean vesicular density			
			Luminal type I vesicles (/μm)	Luminal type II vesicles (/μm)	Cytoplasmic vesicles (/μm ²)	Abluminal vesicles (/μm)
Controls	40	1180	0.34 ± 0.05	0.14 ± 0.03	2.04 ± 0.06	0.21 ± 0.03
<i>Mfsd2a</i> ^{-/-}	40	2449	0.62 ± 0.03 **	0.38 ± 0.03 ***	4.62 ± 0.30 ***	0.48 ± 0.04 ***

b

Density of HRP-filled vesicles in adult brain endothelium (P90)

Tissue source	No. of endothelial profiles	No. of HRP-filled vesicles	Cytoplasmic HRP ⁺ vesicles (/μm ²)
Controls	15	0	0
<i>Mfsd2a</i> ^{-/-}	15	97	3.35 ± 0.55

a. Quantification of the vesicular density (both total and individual type of vesicles) in E17.5 control and mutant endothelium. Mean vesicular density was calculated from the number of vesicular types per μm of luminal membrane (luminal type I and type II vesicles), per μm² of cytoplasm (cytoplasmic vesicles), and per μm of abluminal membrane (abluminal vesicles). **b.** Quantification of HRP luminal uptake in P90 HRP-injected mice. No HRP-filled vesicles were found in wild-type mice. Data are mean ± s.e.m. from 4 controls and 4 mutants (10 vessels per animal, 2 images at ×12,000 per vessel). ***P* < 0.01, ****P* < 0.001, using Student's *t*-test.

Structure of the AcrAB–TolC multidrug efflux pump

Dijun Du¹, Zhao Wang², Nathan R. James¹, Jarrod E. Voss¹, Ewa Klimont¹, Thelma Ohene-Agyei³, Henrietta Venter⁴, Wah Chiu² & Ben F. Luisi¹

The capacity of numerous bacterial species to tolerate antibiotics and other toxic compounds arises in part from the activity of energy-dependent transporters. In Gram-negative bacteria, many of these transporters form multicomponent ‘pumps’ that span both inner and outer membranes and are driven energetically by a primary or secondary transporter component^{1–7}. A model system for such a pump is the acridine resistance complex of *Escherichia coli*. This pump assembly comprises the outer-membrane channel TolC, the secondary transporter AcrB located in the inner membrane, and the periplasmic AcrA, which bridges these two integral membrane proteins. The AcrAB–TolC efflux pump is able to transport vectorially a diverse array of compounds with little chemical similarity, thus conferring resistance to a broad spectrum of antibiotics. Homologous complexes are found in many Gram-negative species, including in animal and plant pathogens. Crystal structures are available for the individual components of the pump^{2–7} and have provided insights into substrate recognition, energy coupling and the transduction of conformational changes associated with the transport process. However, how the subunits are organized in the pump, their stoichiometry and the details of their interactions are not known. Here we present the pseudo-atomic structure of a complete multidrug efflux pump in complex with a modulatory protein partner⁸ from *E. coli*. The model defines the quaternary organization of the pump, identifies key domain interactions, and suggests a cooperative process for channel assembly and opening. These findings illuminate the basis for drug resistance in numerous pathogenic bacterial species.

A small, 49-residue protein, AcrZ, has recently been identified in genetic screens as a binding partner of AcrB that affects substrate preference⁸. A mutant strain of *E. coli* lacking AcrZ is sensitive to some, but not all, antibiotics that are exported by the AcrAB–TolC pump. AcrZ is conserved in many enterobacterial lineages, suggesting that modulation by small proteins may be a recurring theme in the RND protein family to which AcrB belongs.

To investigate their interaction, AcrB was co-expressed with AcrZ carrying a carboxy-terminal histidine tag. The partner proteins co-purified through Ni²⁺ affinity and size-exclusion chromatography (Extended Data Fig. 1a), and co-crystals of the AcrBZ complex were obtained. Additionally, co-crystals were obtained of a ternary complex formed with a designed ankyrin-repeat protein (DARPin)—used in previous studies for co-crystallization with AcrB⁹—that diffracted better than co-crystals of the AcrBZ complex alone (Extended Data Fig. 1b and Extended Data Table 1). Electron density corresponding to AcrZ was apparent in the early, unbiased maps (Extended Data Fig. 2). AcrZ folds into a long, predominantly hydrophobic α -helix that fits into a wide groove in the transmembrane domain of AcrB (Fig. 1a, b and Extended Data Fig. 2). The AcrBZ complex overlays well with the previously reported structure of AcrB bound to another potential partner, YajC¹⁰ (Extended Data Fig. 3a). The helical axis of AcrZ is inclined by nearly 45° with respect to the normal vector of the lipid bilayer, and although this is an unusual extent of inclination for a

transmembrane α -helix, it optimizes the interfacial complementarity. The interfacial residues are well conserved among the identified homologues of AcrZ, suggesting that analogous interactions are likely to occur for other RND family proteins (Extended Data Fig. 3b, c). We were able to model AcrZ residues 1–46, revealing that the amino terminus is in the periplasm and the C terminus is in the cytoplasm, where it interacts with the hydrophilic surface of AcrB, near the membrane (Fig. 1b). It seems possible that the reported effects of AcrZ on drug sensitivity⁸ might originate from the allosteric modulation of AcrB activity by AcrZ.

To stabilize the complete pump assembly, constructs were prepared in which the components were fused (Extended Data Fig. 4a, b). On the basis of the crystal structure of the AcrBZ complex, a construct was designed in which AcrA is fused through a flexible linker to the N terminus of AcrZ, which was predicted to leave the individual structures unperturbed and to permit localization of AcrA to the periplasm while allowing AcrZ to cross the inner membrane with the correct topology and preserving its interactions with AcrB. This fusion harbours a histidine tag on the C terminus of AcrZ to facilitate Ni²⁺ affinity purification of the fusion protein. In a second engineered construct, we identified a loop in the periplasmic domain of AcrB where AcrA could be inserted with flexible linkers. Both the AcrA–AcrZ–His₅ and AcrA–AcrB fusions were stable during overexpression and purification.

At present, there is debate about the compositional stoichiometry of the AcrAB–TolC pump, with evidence favouring a 3:6:3 ratio for AcrB: AcrA: TolC^{11–16}. To enable the assembly of such a complex, we explored the co-expression of the fusion pairs AcrA–AcrB and AcrA–AcrZ–His₅. In the course of isolating the co-expressed recombinant proteins, we observed that the AcrA–AcrZ–His₅ fusion co-purified with both the AcrA–AcrB fusion and endogenous TolC. We subsequently designed co-expression vectors comprising fusion proteins of AcrA–AcrB, AcrA–AcrZ–His₅ and TolC, and found that the proteins remained associated during purification (Extended Data Fig. 4c). Assays of minimum inhibitory concentration (MIC) and drug efflux using the AcrAB and AcrAZ fusion proteins show that the engineered pump retains partial activity (Extended Data Fig. 5).

The assembly was further purified by the GraFix method¹⁷, and analysis of the fractions by negative staining electron microscopy readily led to the identification of drumstick-shaped particles that matched the expected size and shape of the pump (Extended Data Fig. 6a). The crosslinked particles were subsequently imaged using cryo-electron microscopy (cryo-EM), which shows images of isolated and aggregated particles (Fig. 2a). Initial three-dimensional reconstruction was carried out using EMAN2 (ref. 18) without any imposed symmetry (Extended Data Fig. 6b), which suggested the presence of at least three-fold symmetry. As TolC and AcrB are known to be homotrimers, a subsequent three-fold symmetry enforced map was independently produced (Extended Data Fig. 6c). We then used RELION software¹⁹ to refine the particle classification and orientation determination, which improved the map greatly. The resolution of the final map was estimated to be ~16 Å (Extended Data Fig. 7a). This map was validated by a tilt-pair analysis (Extended Data Fig. 7b), which is

¹Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK. ²National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ³Department of Pharmacology, Tennis Court Road, Cambridge CB2 1PD, UK. ⁴School of Pharmacy & Medical Sciences, Sansom Institute for Health Research, University of South Australia, Adelaide, South Australia 5000, Australia.

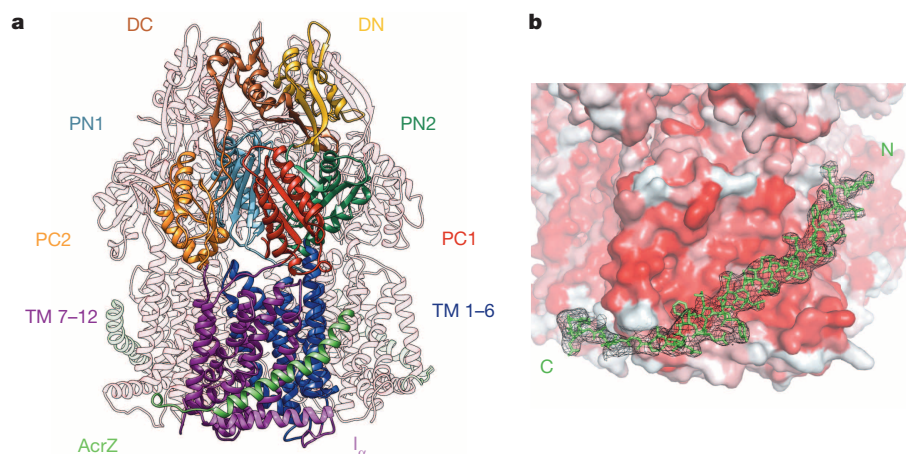


Figure 1 | The crystal structure of the inner-membrane protein AcrB in complex with AcrZ. **a**, Overall view of the AcrBZ trimer. The subdomains of an AcrB protomer are labelled PN1, PN2, PC1, PC2, DN and DC; the transmembrane helices are labelled TM1–6 and TM7–12; and the linking helix is labelled I_{α} . AcrB contains a structural repeat and the structural units comprise PN1–PN2–DN–TM1–6 and PC1–PC2–DC–TM7–12. The protein was crystallized with DARPIn, which is engaged to the AcrB protomer in the crystallographic asymmetric unit (not shown for clarity). **b**, Electron density map with coefficients $2F_o - F_c$ overlaid on the model for AcrZ (green sticks); the contour level is 1σ . AcrB is shown in surface representation, with residues coloured according to their relative hydrophobicity (hydrophobic in red, hydrophilic in white).

considered to be best practice for assessing the correctness of a low-resolution cryo-EM map²⁰.

The quality of the final cryo-EM map was sufficient to permit docking of the crystal structures of TolC, AcrA and the AcrBZ complex with only small movements in the flexible linker regions between the domains of AcrA (Fig. 2b and Methods). After sequential fitting of AcrA, AcrBZ and TolC into the cryo-EM map, the final model has good agreement with the density envelope (Fig. 2b, c and Extended Data Fig. 8a). The transmembrane portion of TolC is enclosed in a torus that resembles the detergent shell observed for aquaporin in the same detergent used in this study, namely *n*-dodecyl- β -D-maltopyranoside (DDM)²¹. The detergent also forms a layer around the transmembrane portion of AcrB (Fig. 2b, c).

One salient observation of our model is that there is no direct interaction between TolC and AcrB in the assembly, in contradiction to earlier models^{3,22}. Instead, the two proteins are bridged in the periplasm entirely through AcrA. This is consistent with *in vivo* crosslinking and pull-down assays, which indicated that AcrA interacts with AcrB and TolC independently, whereas no interaction was observed between AcrB and TolC²³. Our model also accounts for thermodynamic measurements, which suggest that the pairwise interactions of AcrA–AcrB and AcrA–TolC occur spontaneously, whereas that between AcrB and TolC does not²³. The organization of the AcrA–TolC subassembly is in agreement with the dimensions of the homologous MexA–OprM complex from *Pseudomonas aeruginosa*^{24,25}.

Our model shows that the assembly comprises an AcrB trimer, an AcrA hexamer and a TolC trimer. This is in accord with the 3:6 stoichiometry of the homologous CusA–CusB heavy-metal efflux complex¹⁵. The hexameric assembly of AcrA resembles that of the homologous

MacA¹⁶ (Fig. 3a, b). The α -helical coiled coils (or ‘hairpins’) of AcrA pack into a cylinder that interacts with the periplasmic ends of the α -helical coiled coils portion of TolC. As TolC has an internal structural repeat, there are six quasi-equivalent contact surfaces that interact with the AcrA hairpins (Fig. 3a and Extended Data Fig. 8b).

The membrane-proximal domain and β -barrel domain of AcrA are involved in defined interactions with AcrB (Fig. 3a). One protomer of AcrA bridges the upper regions of subdomains PC1, PC2 and DC of AcrB (protomer 1 in Fig. 3a); this interaction is similar to that between CusA and CusB (Fig. 3c). However, the adjacent protomer of AcrA (protomer 2 in Fig. 3a) interacts with AcrB in a different manner: here, the membrane-proximal domain of AcrA shifts towards the PN2 subdomain of AcrB so that only the upper regions of PN2 and DN contact AcrA. The lipoyl domains of AcrA principally interact with each other to form a channel that is closed to the periplasm and make no predicted interactions with either AcrB or TolC.

The side-by-side packing of the β -barrel, lipoyl and α -helical hairpin domains of the six AcrA protomers generates a funnel-like structure with a sealed central channel along the three-fold axis. A chamber is formed by the β -barrel and lipoyl domains of AcrA, and the bottom of the chamber opens into the funnel of AcrB. The chamber is partially constricted near the middle of the lipoyl domain (Fig. 2c and Extended Data Fig. 8a). It then continues to a channel created by the helical hairpin domains, with a cogwheel shape at the end of these domains, like that of MacA (Fig. 3a, b). The modelled docking of the membrane-proximal domain of AcrA onto the surface of AcrB is consistent with the disruptive behaviour of mutants at that interface^{26,27}.

A cross-section through the reconstruction reveals a conduit that is open to the cell exterior and runs from the transmembrane β -barrel of

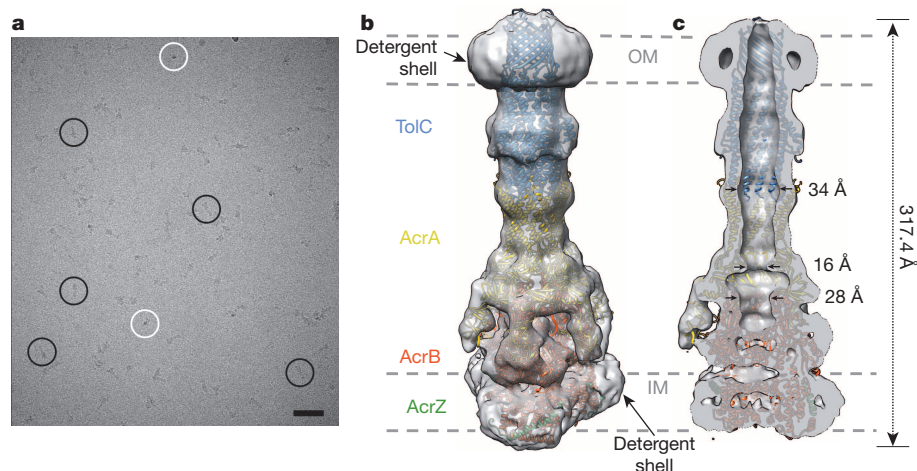


Figure 2 | Cryo-EM images and a pseudo-atomic model of the drug efflux pump. **a**, A representative raw image of the purified pump. White circles indicate particles with long axis almost normal to the viewing plane; black circles show particles with the long axis parallel to the viewing plane. Scale bar, 50 nm. **b**, The reconstructed map and pseudo-atomic model. TolC and AcrBZ are homotrimers that engage six protomers of AcrA to form an assembly with a protein mass of 771 kDa. AcrA has a composite structure of four linearly discontinuous domains connected by flexible linkers⁴ (see also Fig. 3a). **c**, A slice through the reconstruction and model, which shows the continuous conduit that runs from the AcrB funnel through the TolC porin domain. IM, inner membrane; OM, outer membrane.

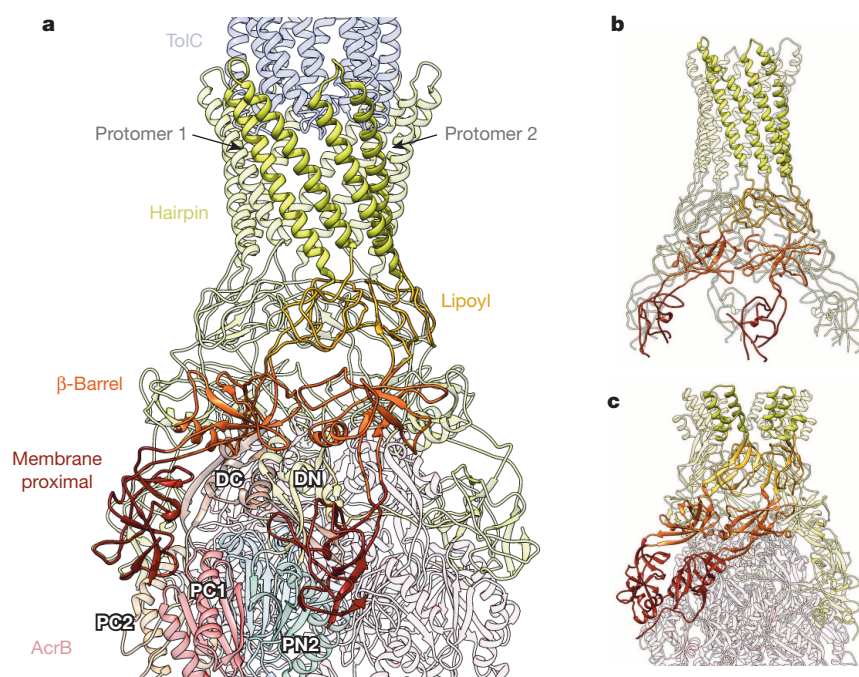


Figure 3 | Modelled interactions of the efflux pump components. **a**, Interactions of the AcrA hairpin domain with TolC and the AcrA β-barrel and membrane-proximal domains with AcrB. The lipoyl domains principally interact with each other and make no interactions with AcrB or TolC. Two AcrA protomers in the homohexameric ring are shown in colour for the hairpin, lipoyl, β-barrel and membrane-proximal domains to illustrate the domain–domain interactions. Subdomains of one protomer of the AcrB trimer are labelled with white labels. The membrane-proximal domains of the two contiguous AcrA protomers make different contacts with the same protomer of AcrB. **b**, Homo-hexameric subunit organization in the structure of MacA, a homologue of AcrA from *Actinobacillus actinomycetemcomitans* (Protein Data Bank (PDB) accession 4DK0). **c**, Interactions between CusA and CusB in the metal-efflux pump from *E. coli* ¹⁵.

TolC, through the TolC periplasmic domain, and continues all the way to the funnel-shaped canyon in the periplasmic domain of AcrB (Fig. 2c). The cross-section indicates that the channel is sealed off from the periplasm and cytoplasm, and the components in the pseudo-atomic model shown in Fig. 2b pack efficiently to leave no gaps. The conduit is the likely exit pathway for efflux of substrates, which will be delivered into this chamber through successive opening and closing of ligand-binding pockets in AcrB, adjacent to the apex of the funnel-shaped periplasmic canyon.

In isolation, TolC assumes a ‘closed’ resting state in which its periplasmic domain tapers to a near close, but to be accommodated in the continuous channel seen in the cryo-EM model, TolC must switch to an ‘open’ state^{28,29} (Fig. 2c and Extended Data Fig. 8b). The opening of TolC must result from its direct interactions with the α-helical hairpins of AcrA. In this scenario, AcrB provides a platform for the proper assembly of hexameric AcrA suitable for presentation to TolC and opening of the channel. The AcrAB–TolC assembly forms *in vivo* without the requirement for substrates or proton-motive force³⁰, suggesting that exogenous energy may not be required to open TolC in the context of the assembled pump. It seems likely that TolC remains open throughout the transport cycle, and we propose that the dilation of TolC from the closed to the open state is driven by both chelate cooperativity (due to the hexameric organization of AcrA) and allosteric cooperativity (associated with the breaking of interprotomer interactions in TolC).

The structure of the efflux pump presented here is likely to be similar to homologous assemblies in pathogenic species that infect humans such as *Vibrio cholerae*, *P. aeruginosa*, *Neisseria gonorrhoeae* and *Salmonella enterica*. The question arises as to how it might be feasible to counter the actions of drug efflux in the treatment of bacterial infections. It seems paradoxical that a compound might exist that would achieve this, when the pump is well crafted to handle diverse compounds. Perhaps the protein–protein contacts that mediate the assembly might be favourable targets for such approaches.

METHODS SUMMARY

Recombinant proteins were engineered with histidine tags, overexpressed in *E. coli* and purified chromatographically. The AcrBZ and AcrBZ–DARPin complexes were crystallized by vapour diffusion and diffraction data were collected at the Diamond Light Source synchrotron facility. The crystal structure was solved using molecular replacement and refined. Freshly prepared specimens of the AcrABZ–TolC complex

were applied to glycerol–glutaraldehyde gradients with ultracentrifugation¹⁷. The fractions were analysed by electron microscopy after either negative staining or rapid freezing under hydrated conditions. Individual particles were chosen from the cryo-EM images and were used for two-dimensional projection class averages and three-dimensional image reconstruction^{18,19}. For *in vivo* activity tests of the engineered AcrABZ–TolC pump, expression vectors for the fusion proteins and controls were transformed into deletion strains. MICs were evaluated in the presence of antibiotics and efflux assays were conducted on fluorescent substrates.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 September 2013; accepted 5 March 2014.

Published online 20 April 2014.

- Du, D., Venter, H., Pos, K. M. & Luisi, B. F. in *Microbial Efflux Pumps: Current Research* Ch. 3 (Caister Academic, 2013).
- Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. F. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).
- Murakami, S., Nakashima, R., Yamashita, E. Cry & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).
- Mikolosko, J., Bobyk, K., Zgurskaya, H. I. & Ghosh, P. Conformational flexibility in the multidrug efflux system protein AcrA. *Structure* **14**, 577–587 (2006).
- Seeger, M. A. *et al.* Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **313**, 1295–1298 (2006).
- Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443**, 173–179 (2006).
- Eicher, T. *et al.* Transport of drugs by the multidrug transporter AcrB involves an access and a deep binding pocket that are separated by a switch-loop. *Proc. Natl Acad. Sci. USA* **109**, 5687–5692 (2012).
- Hobbs, E. C., Yin, X., Paul, B. J., Astarita, J. L. & Storz, G. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proc. Natl Acad. Sci. USA* **109**, 16696–16701 (2012).
- Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M. G. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol.* **5**, e7 (2007).
- Törnroth-Horsefield, S. *et al.* Crystal structure of AcrB in complex with a single transmembrane subunit reveals another twist. *Structure* **15**, 1663–1673 (2007).
- Janganan, T. K. *et al.* Evidence for the assembly of a bacterial tripartite multidrug pump with a stoichiometry of 3:6:3. *J. Biol. Chem.* **286**, 26900–26912 (2011).
- Stegmeier, J. F., Polleichtner, G., Brandes, N., Hotz, C. & Andersen, C. Importance of the adaptor (membrane fusion) protein hairpin domain for the functionality of multidrug efflux pumps. *Biochemistry* **45**, 10303–10312 (2006).
- Narita, S.-i., Eda, S., Yoshihara, E. & Nakae, T. Linkage of the efflux-pump expression level with substrate extrusion rate in the MexAB–OprM efflux pump of *Pseudomonas aeruginosa*. *Biochem. Biophys. Res. Commun.* **308**, 922–926 (2003).
- Mima, T., Joshi, S., Gomez-Escalada, M. & Schweizer, H. P. Identification and characterization of TriABC–OpmH, a triclosan efflux pump of *Pseudomonas*

- aeruginosa* requiring two membrane fusion proteins. *J. Bacteriol.* **189**, 7600–7609 (2007).
15. Su, C. C. *et al.* Crystal structure of the CusBA heavy metal efflux complex of *Escherichia coli*. *Nature* **470**, 558–562 (2011).
 16. Yum, S. *et al.* Crystal structure of the periplasmic component of a tripartite macrolide-specific efflux pump. *J. Mol. Biol.* **387**, 1286–1297 (2009).
 17. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature Methods* **5**, 53–55 (2008).
 18. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
 19. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
 20. Henderson, R. *et al.* Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J. Mol. Biol.* **413**, 1028–1046 (2011).
 21. Berthaud, A., Manzi, J., Pérez, J. & Mangenot, S. Modeling detergent organization around aquaporin-O using small angle X-ray scattering. *J. Am. Chem. Soc.* **134**, 10080–10088 (2012).
 22. Symmons, M. F., Bokma, E., Koronakis, E., Hughes, C. & Koronakis, V. The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proc. Natl Acad. Sci. USA* **106**, 7173–7178 (2009).
 23. Touzé, T. *et al.* Interactions underlying assembly of the *Escherichia coli* AcrAB–TolC multidrug efflux system. *Mol. Microbiol.* **53**, 697–706 (2004).
 24. Xu, Y. *et al.* Assembly and channel opening of outer membrane protein in tripartite drug efflux pumps of Gram-negative bacteria. *J. Biol. Chem.* **287**, 11740–11750 (2012).
 25. Trépout, S. *et al.* Structure of reconstituted bacterial membrane efflux pump by cryo-electron tomography. *Biochim. Biophys. Acta* **1798**, 1953–1960 (2010).
 26. Tikhonova, E. B., Yamada, Y. & Zgurskaya, H. I. Sequential mechanism of assembly of multidrug efflux pump AcrAB–TolC. *Chem. Biol.* **18**, 454–463 (2011).
 27. Ge, Q., Yamada, Y. & Zgurskaya, H. The C-terminal domain of AcrA is essential for the assembly and function of the multidrug efflux pump AcrAB–TolC. *J. Bacteriol.* **191**, 4365–4371 (2009).
 28. Janganan, T. K., Bavro, V. N., Zhang, L., Borges-Walmsley, M. I. & Walmsley, A. R. Tripartite efflux pumps: energy is required for dissociation, but not assembly or opening of the outer membrane channel of the pump. *Mol. Microbiol.* **88**, 590–602 (2013).
 29. Bavro, V. N. *et al.* Assembly and channel opening in a bacterial drug efflux machine. *Mol. Cell* **30**, 114–121 (2008).
 30. Tikhonova, E. B. & Zgurskaya, H. I. AcrA, AcrB, and TolC of *Escherichia coli* form a stable intermembrane multidrug efflux complex. *J. Biol. Chem.* **279**, 32116–32124 (2004).

Acknowledgements This work was supported by the Wellcome Trust and Human Frontier Science Program (B.L.) and partly by a National Institutes of Health grant (P41GM103832 to W.C.). J.E.V. is supported by a Herschel Smith scholarship. T.O.-A. is the recipient of a Cambridge Trust scholarship, an Adam Glinzman award and a Faculty for the Future Fellowship from the Schlumberger Foundation. We thank M. Pos for kindly providing pBAD-AcrAB-TolC. We thank L. Packman for mass-spectrometric analyses and S. J. Ludtke, M. F. Schmid, M. Pos and R. van Veen for helpful advice and discussions. We are grateful to the staff at the Diamond Light Source for access to facilities and invaluable help.

Author Contributions D.D., W.C. and B.F.L. designed the experiments; D.D. and N.R.J. purified and crystallized AcrBZ complexes. D.D. and B.F.L. solved the crystal structure of AcrBZ complexes. D.D., N.R.J. and E.K. purified AcrABZ–TolC complexes. Z.W., J.E.V. and W.C. obtained and analysed the single-particle cryo-EM data. D.D. and B.F.L. devised a model of AcrABZ–TolC based on the cryo-EM map. T.O.-A. and H.V. conducted MIC and efflux assays on the AcrABZ–TolC pump. D.D., J.E.V., W.C. and B.F.L. analysed results. All authors contributed to writing the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the PDB under accessions 4C48 (AcrB/AcrZ/DARPin) and 4CDI (AcrB/AcrZ). The cryo-EM map has been deposited in the Electron Microscopy Data Bank under accession EMD-5915. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.F.L. (bfl20@cam.ac.uk) or D.D. (dd339@cam.ac.uk).

METHODS

Construction of vectors for overexpression of AcrBZ and AcrABZ–TolC complex.

The *acrZ* gene was generated by PCR from genomic DNA of *E. coli* W3110 strain using primers AcrZ_{NcoI_F}: 5'-TGGCCATGGGCTTAGAGTTATTAAGTCTGGTATTCGCCGTAATCATGG-3' and AcrZ_{His5_Sall_R}: 5'-GACGTCGACTCAGTGGTGGTGGTGGTGGTATGATTTTGTCCGGGCTGGTCTTTTACC-3'. The pRSFDuet-1-*acrZ*_{His5} vector was constructed by subcloning the fragment of *acrZ*-*His5*, bounded by NcoI and Sall sites, into the multiple cloning site (MCS) of expression vector pRSFDuet-1.

The pET21a-*acrB*_{His6} vector was constructed by subcloning the fragment of *acrB*, bounded by NdeI and XhoI sites, into the MCS of expression vector pET21a. The vector pET21a-*acrB*_{His6} was constructed by site-directed mutagenesis using pET21a-*acrB*_{His6} as a template and primers AcrB_CF: 5'-TGAAGATATCGAGCACAGC CATACTGTCGATTGAGATCCGGCTGCTAACAAGCCC-3' and AcrB_CR: 5'-GGGCTTTGTTAGCAGCCGGATCTCAATCGACAGTATGGCTGTGCTC GATATCTTCA-3'.

The *acrA* gene was amplified using primers AcrANcoI_F: 5'-TGGCCATGG GCAACAAAAACAGAGGGTTACGCTCTGGCG-3' and AcrAGSx3_R: 5'-T GGCCATGGATCCGCCGCCACGACGCCACCCGCGCTCCACCGCCA CCAGACTTGGACTGTTCAGGCTGAGCACC-3', to add a poly-GlySer linker to the C terminus and a NcoI site at both ends. The NcoI-bounded *acrA*-polyGlySer was subcloned into the NcoI site of pRSFDuet-1-*acrZ*_{His5}, resulting in the construct pRSFDuet-1-*acrA*-polyGlySer-*acrZ*_{His5}. The *tolC* gene was amplified using primers TolCinf_F: 5'-AAGGAGATATACATATGAAGAAATTGCTCCCCATTCTTA TCGGCC-3' and TolC1392inf_R: 5'-TTGAGATCTGCCATATGTCAATCAG CAATAGCATTCTGTTCGGCGCT-3'. The PCR product was then inserted into the NdeI site of pRSFDuet-1-*acrA*-polyGlySer-*acrZ*_{His5} using the In-Fusion cloning method (Clontech), generating the construct pRSFDuet-1-*acrA*-polyGlySer-*acrZ*_{His5}-*tolC*₁₃₉₂.

A BamHI site was inserted into *acrB* by site-directed mutagenesis using pET21a-*acrB*_{His6} as a template and primers AcrB_D328F: 5'-CTGAAAATTGTT TACCCATAGACGGATCCACCACGCCGTTTCGTGAAA-3' and AcrB_D328R: 5'-TTTCCAGAACGGCGTGGTGGATCCGTCGTATGGGTAACAAATTTTCA G-3', resulting in the construct pET21a-*acrB*_{His6}BamHI₃₂₈. The *acrA* gene was amplified using primers AcrA_25GSF: 5'-GGATCCGGTGGGAGCGGTGGCGG CGGTAGTGGCGGTGGTGGCTCTTGTGACGACAAACAGGCCAACAGG-3' and AcrAGSx3_R2: 5'-GGATCCGCCGCCACAGAGCCACCACCGCGCTCCCA CCGCCACAGACTTGGACTGTTCAGGCTGAGCACC-3', to add a poly-GlySer linker to both ends of *acrA*. The resulting DNA fragment was amplified again using primers AcrA Infusion_F: 5'-CCCATACGACGGATCCGGTGGGAGCG GTGGCGCGCGTA-3' and AcrA Infusion_Reverse: 5'-ACGCGTGGTGGAT CCGCCGCCACAGAGCCACCACC-3'. The DNA fragment was then inserted into the BamHI site of pET21a-*acrB*_{His6}BamHI₃₂₈ using the In-Fusion cloning method, generating the construct pET21a-*acrB*₃₂₈-polyGlySer-*acrA*-polyGlySer-*acrB*₃₂₉.

Overexpression and purification of the AcrBZ and the AcrBZ–DARPin complexes. The constructs pET21a-*acrB*_{His6} and pRSFDuet-1-*acrZ*_{His5} were transformed into *E. coli* strain C43(DE3) Δ acrAB³¹. The C-terminal histidines of AcrB were substituted to prevent nonspecific association with the metal affinity matrix. Cells were grown in 2 \times YT medium with 100 μ g ml⁻¹ carbenicillin and 50 μ g ml⁻¹ kanamycin at 37 °C until the culture reached an absorbance, at 600 nm, of 0.5–0.6 and was then induced by the addition of 0.5 mM isopropyl 1-thio- β -D-galactopyranoside (IPTG) at 18 °C overnight. Cell pellets were resuspended in lysis buffer (400 mM sodium chloride, 20 mM Tris-HCl, pH 8.0) with 1 tablet per 50 ml EDTA-free protease inhibitor cocktail tablets, 5 U ml⁻¹ DNase I and 5 mg ml⁻¹ lysozyme, and the mixture was stirred at 4 °C for 1 h to digest the cell wall. The cells were lysed using a homogenizer (EmulsiFlex) at 15,000 psi for eight passages. Cell debris was pelleted by centrifugation at 9,000g for 30 min. Cellular membrane was pelleted by ultracentrifugation at 125,755g for 3 h.

Membrane pellets were resuspended in lysis buffer with protease inhibitors and were solubilized by adding 1.5% DDM and stirring at 4 °C for 3 h. Debris was pelleted by ultracentrifugation at 125,755g for 30 min. Imidazole was added to the membrane solution to a final concentration of 10 mM. Histidine-tagged AcrBZ complex was purified by nickel affinity chromatography using a HiTrap chelating column (GE Healthcare Life Sciences) equilibrated with GF buffer 1 (400 mM sodium chloride, 20 mM Tris-HCl, pH 8.0, 0.05% DDM) containing 20 mM imidazole. The column was washed with 50 mM and 75 mM imidazole added to GF buffer 1, respectively. Purified AcrBZ complex was eluted with 500 mM imidazole in GF buffer 1, concentrated and loaded onto a Superose 6 column equilibrated with GF buffer 1. Fractions containing purified AcrBZ complex were pooled and concentrated to 15–20 mg ml⁻¹ using a Vivaspin concentrator (MWCO = 100 kDa) and dialysed overnight against sample buffer (10 mM HEPES pH 7.5, 50 mM sodium chloride, 0.03% DDM) using a 100 kDa MWCO dialysis membrane to decrease the detergent concentration.

The DARPin gene was synthesized and protein was overexpressed and purified as described⁹. Purified DARPin and AcrBZ complex were mixed at a molar ratio of 1:2 (AcrBZ monomer:DARPin monomer). The mixture was diluted with GF buffer 2 (20 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 0.03% DDM) to a concentration of 2–3 mg ml⁻¹, incubated at 4 °C overnight, then concentrated to 0.5 ml using a Vivaspin concentrator (MWCO = 100 kDa) and loaded onto a Superose 6 column equilibrated with GF buffer 2. Fractions containing purified AcrBZ–DARPin complex were pooled and concentrated to 15–20 mg ml⁻¹ using a Vivaspin concentrator (MWCO = 100 kDa) and dialysed overnight against sample buffer using a 100 kDa MWCO dialysis membrane; the final concentration was 10–15 mg ml⁻¹.

Crystallization of AcrBZ and AcrBZ–DARPin. The AcrBZ and AcrBZ–DARPin complexes were diluted to 10 mg ml⁻¹ using sample buffer. 9 mM n-octyl- β -D-thiogluco-pyranoside (90 mM) was mixed with the AcrBZ complex before the crystallization trials. The AcrBZ crystals were grown at 20 °C using the hanging-droplet vapour diffusion method by mixing 4 μ l of AcrBZ complex with 2 μ l of reservoir solution (100 mM tricine, pH 7.4, 50 mM lithium sulphate, 5 mM cadmium chloride hydrate, 7% PEG 3000, 10% glycerol). The AcrBZ–DARPin complex was incubated with puromycin at 1 mM for 3 h at 4 °C before crystallization trials. Crystals were grown at 20 °C using the sitting-droplet vapour diffusion method by mixing 400 nl of AcrBZ–DARPin complex with 200 nl of reservoir solution (100 mM HEPES, pH 7.5, 10 mM MgCl₂ and 12% (w/v) PEG 3350). Crystals appeared 24 h after setting up the crystallization trials and reached maximal size in 1 week. The crystals were transferred briefly into reservoir solution supplemented with 25% glycerol as cryoprotectant before flash freezing in liquid nitrogen.

Crystallographic data collection and structure refinement. Data sets were collected using beamlines I04-1 and I24 at the Diamond Light Source. The diffraction data were processed using X-ray Detector Software (XDS)³² and scaled using SCALA³³. Structures were solved by molecular replacement using Phaser with chain A of AcrB (PDB accession 4DX5)⁷ as a search model and refined using Crystallography and NMR System (CNS)^{34,35}, PHENIX³⁶ and REFMAC5 (ref. 37) with reference structure restraints³⁸. Coot was used for modelling³⁹. Data collection and refinement parameters are presented in Extended Data Table 1.

Both the AcrBZ and AcrBZ–DARPin complexes crystallized in space group R32. The asymmetric unit is occupied by one protomer of AcrB and one protomer of AcrZ. Maps calculated from molecular replacement using only the AcrB protomer revealed clear electron density for AcrZ in both complexes (see Extended Data Fig. 2). For the AcrBZ–DARPin complex, unbiased density was also present that could fit the DARPin model. Density was also apparent that matched the modelled detergent and acyl chains in the 1.9 Å resolution structure of *E. coli* AcrB⁷ (PDB accession 4DX5). These were included in the model, as well as additional acyl chains that were located mostly in the intra-protomer space in the transmembrane region of AcrB. For both AcrBZ and AcrBZ–DARPin complexes, the AcrB homotrimer is formed through crystallographic symmetry, so that any potential non-equivalence of the protomers is consequently lost.

Expression and purification of the efflux pump assembly. The constructs pET21a-*acrB*₃₂₈-polyGlySer-*acrA*-polyGlySer-*acrB*₃₂₉ and pRSFDuet-1-*acrA*-polyGlySer-*acrZ*_{His5}-*tolC*₁₃₉₂ were transformed into *E. coli* strain C43(DE3) Δ acrAB. The procedure for overexpression and purification of AcrABZ–TolC is similar to that of AcrBZ, with minor modifications, as follows. Cells were induced with 0.15 mM IPTG at 20 °C overnight, harvested and resuspended in lysis buffer 2 (20 mM Tris-HCl, pH 7.5, 400 mM NaCl) for membrane preparation. Cellular membrane was solubilized with 1.5% DDM in lysis buffer 2. The AcrABZ–TolC complex purified by nickel affinity chromatography was loaded onto a Superose 6 column equilibrated with GF buffer 3 (50 mM HEPES pH 7.5, 400 mM NaCl, 0.03% DDM). Fractions containing purified AcrABZ–TolC complex were pooled and concentrated to 1 mg ml⁻¹ using a Vivaspin column (MWCO = 100 kDa).

Transport assays. Transport activity was measured as described⁴⁰ with some modifications. Plasmids were propagated in the triple knockout *E. coli* strain MCAtolCAcrAB^{23,41}, a derivative of *E. coli* MC1061 converted to a (DE3) lysogen. All experiments used basal levels of protein expression without induction with IPTG. Cells were grown in LB-Broth Miller medium (Formedium) containing the appropriate antibiotic(s) with shaking at 37 °C until an OD_{600 nm} of 0.5 was reached. The cells were harvested by centrifugation at 3,000g for 10 min at 4 °C, then washed three times by resuspension in 50 mM potassium phosphate buffer, pH 7.0, containing 5 mM MgSO₄ and sedimented by centrifugation at 3,000g for 10 min at 4 °C. The cells were then resuspended in the same buffer to an OD_{600 nm} of 0.5 and incubated for 3 min at room temperature (25 °C) in the presence of 25 mM glucose to energize the cells. The fluorescence measurement was started, and 60 s later 2 μ M ethidium bromide or 0.25 μ M trimethylammonium-diphenylhexatriene (TMA-DPH) was added. The fluorescence was followed as a function of time in a PerkinElmer LS 55B fluorimeter. Excitation and emission wavelengths and excitation and emission slit widths were 500, 580, 5 and 10 nm respectively for ethidium bromide; and 350, 425, 5 and 5 nm respectively for TMA-DPH. Initial substrate

transport rates were determined between 85 to 185 s and 75 to 95 s for ethidium bromide and TMA-DPH, respectively. Assays were performed twice per day for three separate days, with different cell preparations used each day.

GraFix preparation of the efflux pump assembly. The GraFix procedure for the AcrABZ–TolC complex was carried out as described^{17,42}. Gradients for GraFix were generated by mixing GraFix buffer 1 (50 mM HEPES, pH 7.5, 400 mM NaCl, 0.03% w/v DDM, 10% v/v glycerol) and GraFix buffer 2 (50 mM HEPES, pH 7.5, 400 mM NaCl, 0.03% w/v DDM, 30% v/v glycerol, 0.15% v/v glutaraldehyde) using a gradient mixer (Gradient Master 107, BioComp Instruments), following the manufacturer's recommendations for determining the parameters. 100 pmol of AcrABZ–TolC complex was loaded onto the top of the buffering cushion. Ultracentrifugation was carried out at 111,845g, 4 °C for 18 h (SW60 rotor, Beckmann). After centrifugation, the gradients were fractionated at 4 °C by using a capillary to pump the gradient out from bottom to top, taking fractions of 200 μ l. The same fractions from different centrifuge tubes were pooled, respectively. Glycine (1 M, pH 7.5) was added to the fractions to a final concentration of 80 mM to quench further crosslinking, and the mixtures were concentrated to 200 μ l using a Vivaspinn concentrator (MWCO = 100 kDa). The concentrated protein samples were diluted to 1 ml using GF buffer 3, and were concentrated to 100 μ l again to decrease the concentration of glycerol. The AcrABZ–TolC samples were flash frozen in liquid nitrogen and stored at –80 °C.

Negative-stain electron microscopy. Continuous carbon electron microscopy grids (Agar Scientific) were glow discharged. Sample adsorption was achieved by placing the grid carbon-side down on a 10 μ l drop of protein solution at a concentration of approximately 10 μ g ml^{–1}. Adsorption times were determined empirically. After blotting to dryness, the grid was washed by placing it onto a drop of deionized H₂O for 5 s, blotted dry, and then repeated. The protein was stained by transferring the grid to a drop of 2% w/v uranyl acetate for 30 s, then dried and stored at room temperature. The sample was then imaged at $\times 20,000$ magnification and 120 keV in a FEI Tecnai G² microscope at the Multi Imaging Centre, University of Cambridge.

Cryo-EM specimen preparation and data collection. Samples were embedded in vitreous ice as per the following method. A 2.0 μ l aliquot at approximately 2 mg ml^{–1} was applied onto holey carbon film supported by a 200-mesh R1.2/1.3 Quantifoil grid (Quantifoil) that had been previously washed and glow discharged. The grid was blotted and rapidly frozen in liquid ethane using a Vitrobot IV (FEI) with constant temperature and humidity during the process of blotting, and the grid was stored in liquid nitrogen before imaging. All grids were screened on a JEM3200FSC cryo-electron microscope (JEOL) operated at 300 kV, spot size 2, condenser aperture 70 μ m, objective aperture 60 μ m, and energy slit of the in-column filter of 20 eV. Images were recorded with a direct detection device (DDD) (DE-12 3kx4k camera, Direct Electron, LP) operating in movie mode at a recording rate of 25 raw frames per second. A magnification of $\times 20,000$ (corresponding to a calibrated sampling of 2.47 Å per pixel) and a dose rate of 25 electrons s^{–1} Å^{–2} were used to acquire data for each specimen area with a total exposure time of 2 s. A total of 1,281 DDD images using movie frame mode were collected with a defocus range of 2–4 μ m.

Cryo-EM data processing and resolution evaluation. Particle images were manually boxed with the EMAN2 program e2boxer.py¹⁸ with a box size of 256 \times 256 pixels using the averaged sum of 48 raw frames per specimen area. The final frame average was computed from averages of every five consecutive frames to correct for the specimen movement during exposure similar to the procedure previously described⁴³. The particle intensity in each frame was weighted according to a radiation damage model (courtesy of B. Bammes). Defocus of the particles in each frame average was determined by EMAN2 program e2ctf.py⁴⁴. The first reconstruction was done without any symmetry imposition or any initial model, using EMAN2 (ref. 18). We started with 12,000 particle images and 8,400 particle images were selected manually according to the similarity in appearance to the class averages. The first map without any symmetry imposition appeared to have a six-fold symmetry (Extended Data Fig. 6b), which justified applying a three-fold symmetry enforcement (Extended Data Fig. 6c). The symmetry-free and symmetry-imposed maps look similar except for a slight difference in contrast. Then, we used RELION¹⁹ to perform two additional steps of particle orientation refinement of the 8,400 selected particle images. Using the EMAN2-generated map (Extended Data Fig. 6c), low-pass filtered to 60 Å, we performed a particle orientation refinement by RELION with search angle of 7.5 degrees for 14 iterations and subsequently with an angular sampling of 1.8 degrees for 11 iterations. The resolution of the resulting map is 21 Å based on 0.143 FSC criterion (Extended Data Fig. 7a). Next, this map was sharpened¹⁸ and low-pass filtered to 30 Å before performing a second round of particle orientation refinement by RELION with an angular sampling of 1.8 degrees for 8 iterations. The resolution of the final map is estimated to be 16 Å by the 0.143 FSC criterion (Extended Data Fig. 7a). The final map displayed in Fig. 2b, c has been sharpened and filtered to 16 Å.

Cryo-EM map tilt-pair validation. To further validate the cryo-EM map (Fig. 2b), we carried out tilt-pair analysis²⁰ of the AcrABZ–TolC complex. Pairs of images were recorded on a JEM2010F electron microscope at a magnification of $\times 30,000$ with a relative tilt of 10 degrees, along a known axis. Each pair of particle images thus has a known experimental relative tilt. We then determined the orientation of each particle in the pair using the final reconstructed map as a reference. The relative tilt between these computed orientations should match the experimental tilt. Each point in Extended Data Fig. 7b represents one pair of particles in the tilt pair, with radius representing computed relative tilt, and azimuth representing tilt angle. Ideally all points would lie in exactly the same location; however, there is always some uncertainty in orientation determination, exacerbated in this case by radiation damage in the second image of the tilt pair and the strong pseudo hexagonal symmetry. As previously observed⁴⁵, if the map is incorrect, the relative angles will not correlate at all, and produce a nearly random distribution over the sphere. A clear cluster as observed in our case is an indication of successful validation, and our computed tilt angle of 10.44 degrees is a good match with the microscope setting of 10 degrees. The presence of some points at mirrored positions across the origin is an indication of weak handedness in the map, meaning that the map and its mirror image are difficult to distinguish at this resolution in some orientations.

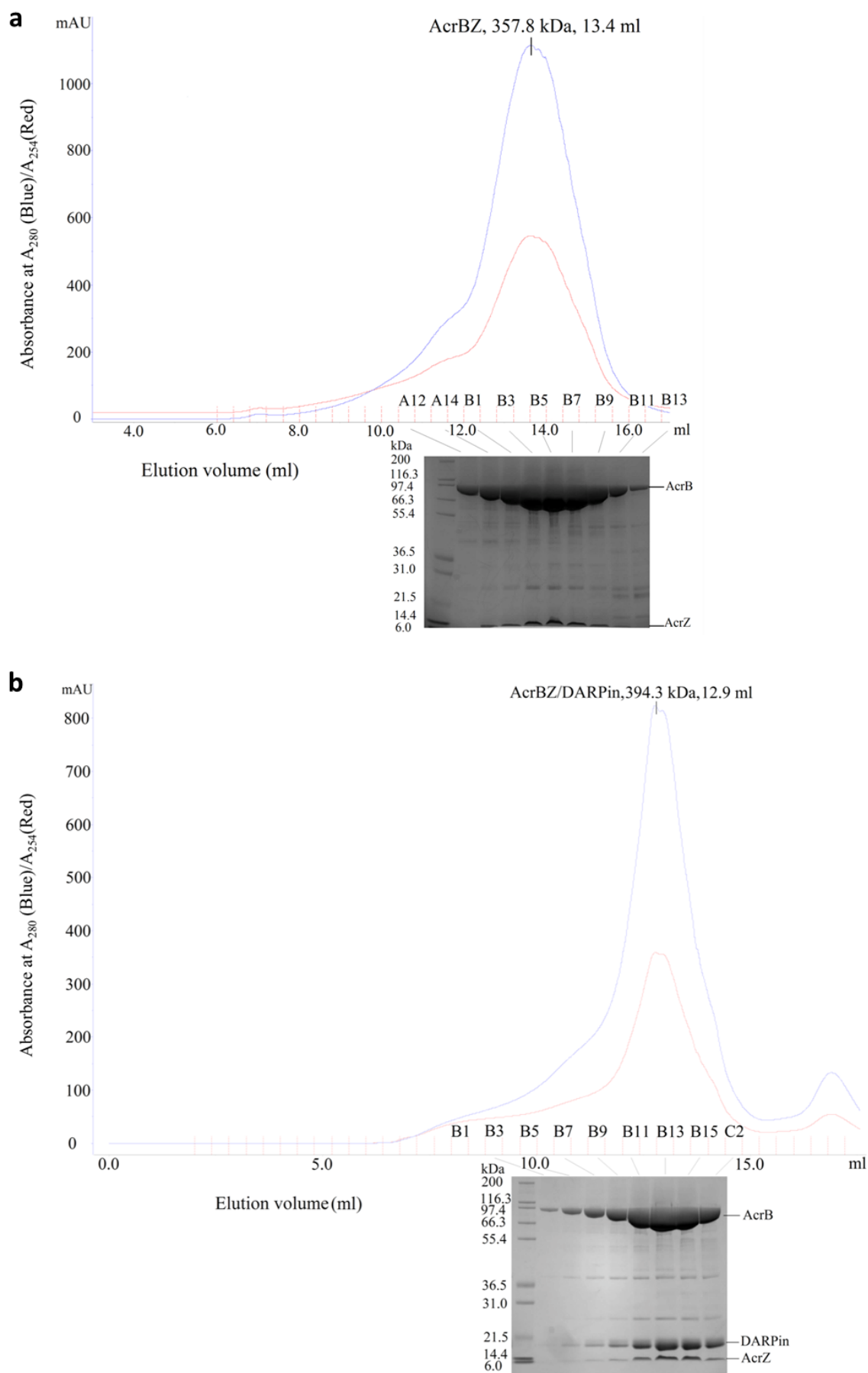
Model docking. The merged map was low-pass filtered before rigid body docking of crystal structures using Coot and UCSF Chimera⁴⁶ (<http://www.cgl.ucsf.edu/chimera>). The symmetric partially opened model of TolC was used to prepare a fully opened model (PDB accession 2XMN)⁴⁷. A homology model of the inner sets of coiled coils was devised based on the outer sets of coiled coils, which was used to replace the inner sets of coiled coils. The uncoiling movement of the inner sets of coiled coils produced a fully opened TolC².

The β -barrel domain, lipoyl domain and α -helical hairpin domain of each AcrA protomer were modelled with chain C in the asymmetric unit of the crystal structure of AcrA (PDB accession 2F1M)⁴. A homology model of the membrane proximal domain of AcrA was devised based on the structure of MexA (PDB accession 2V4D)²².

The individual domains of AcrA were fitted into the cryo-EM reconstruction using Coot, according to the features of the cryo-EM map and prior knowledge of the homohexameric MacA assembly (PDB accessions 3FPP and 4DK0). The closed-state TolC (PDB accession 1EK9), partially opened (PDB accession 2XMN) and fully opened model of TolC were separately fitted into the cryo-EM reconstruction, respectively. The AcrBZ trimer, the TolC trimer and the AcrA hexamer were treated as individual rigid bodies. These three rigid bodies were automatically fitted into the cryo-EM map by sequential fitting using Chimera. The AcrABZ–TolC model generated by Chimera was slightly adjusted manually to optimize the local fit into the density using Coot³⁹. The cross-correlation coefficient for fitting the AcrABZ–open-TolC model to the map is 0.924 and was calculated using the 'fit in map' function of Chimera.

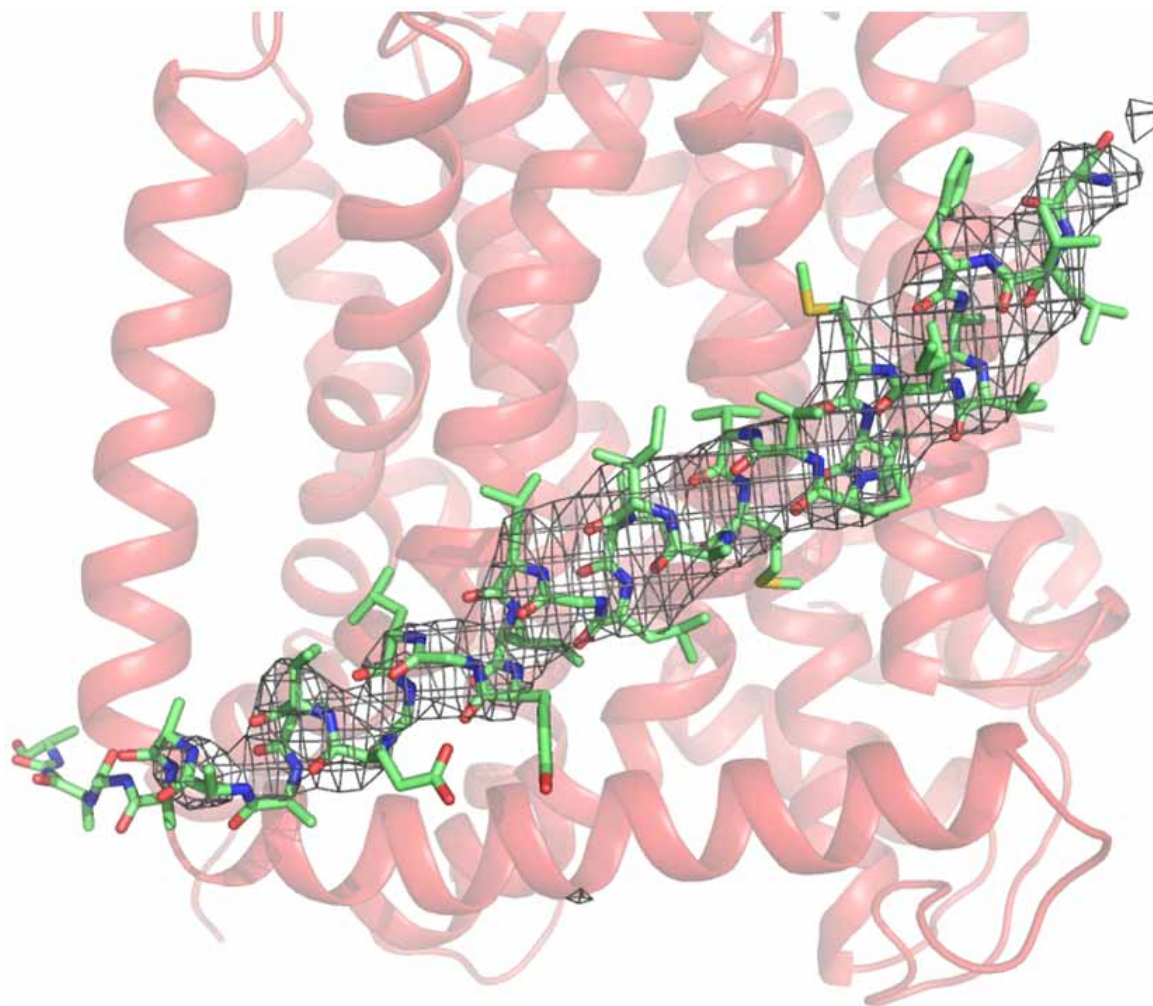
1. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
2. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
3. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
4. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
5. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
6. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
7. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55**, 247–255 (1999).
8. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
9. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
10. Welch, A., Awah, C. U., Jing, S., van Veen, H. W. & Venter, H. Promiscuous partnering and independent activity of MexB, the multidrug transporter protein from *Pseudomonas aeruginosa*. *Biochem. J.* **430**, 355–364 (2010).
11. Lobedanz, S. et al. A periplasmic coiled-coil interface underlying TolC recruitment and the assembly of bacterial drug efflux pumps. *Proc. Natl Acad. Sci. USA* **104**, 4612–4617 (2007).
12. Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol.* **481**, 109–126 (2010).
13. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
14. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
15. Murray, S. C. et al. Validation of cryo-EM structure of IP₃R1 channel. *Structure* **21**, 900–909 (2013).

46. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
47. Pei, X. Y. *et al.* Structures of sequential open states in a symmetrical opening transition of the TolC exit duct. *Proc. Natl Acad. Sci. USA* **108**, 2112–2117 (2011).
48. Tsukazaki, T. *et al.* Structure and function of a membrane component SecDF that enhances protein export. *Nature* **474**, 235–238 (2011).
49. Hsieh, Y. H. *et al.* Reconstitution of functionally efficient SecA-dependent protein-conducting channels: transformation of low-affinity SecA-liposome channels to high-affinity SecA-SecYEG-SecDF-YajC channels. *Biochem. Biophys. Res. Commun.* **431**, 388–392 (2013).
50. du Plessis, D. J., Nouwen, N. & Driessen, A. J. The Sec translocase. *Biochim. Biophys. Acta* **1808**, 851–865 (2011).
51. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
52. Jiang, W. *et al.* Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* **308**, 1033–1044 (2001).



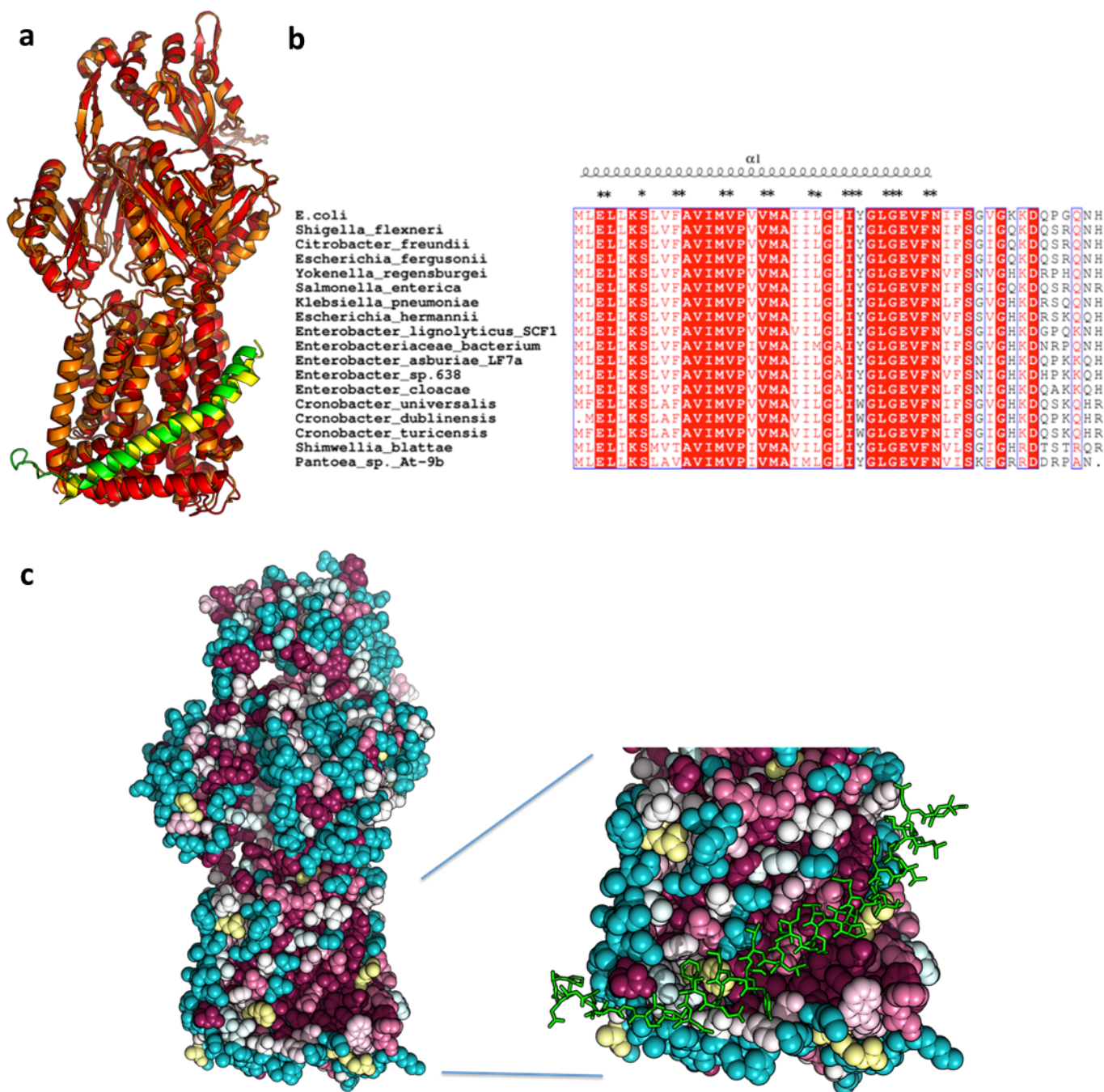
Extended Data Figure 1 | Co-purification of the AcrBZ and AcrBZ-DARPin complexes. **a**, Gel filtration profile of AcrBZ complex and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the peak fractions. **b**, Gel filtration profile of the AcrBZ-DARPin complex and SDS-PAGE analysis. The proteins were enriched by immobilized metal affinity chromatography (IMAC; results not shown) before the size-exclusion

chromatography step. AcrZ has a heptahistidine tag on the C terminus as an IMAC affinity tag, and the C-terminal histidines of AcrB have been removed to prevent its binding to the matrix. Standard proteins thyroglobulin (669 kDa) and aldolase (158 kDa) (Bio-Rad) were eluted from the same column at volumes of 11.4 and 14.7 ml, respectively. See also Fig. 1.



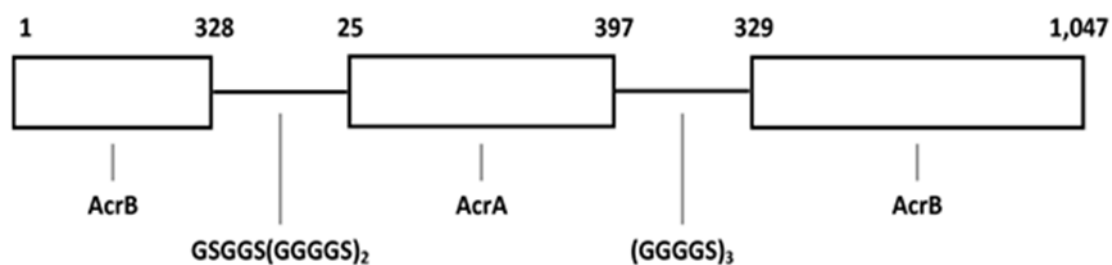
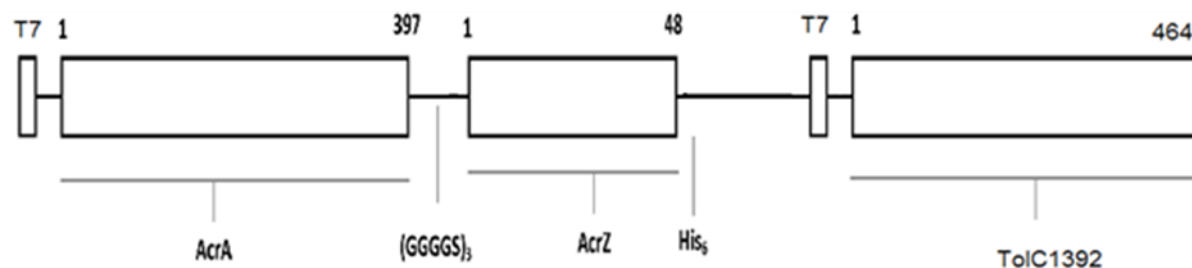
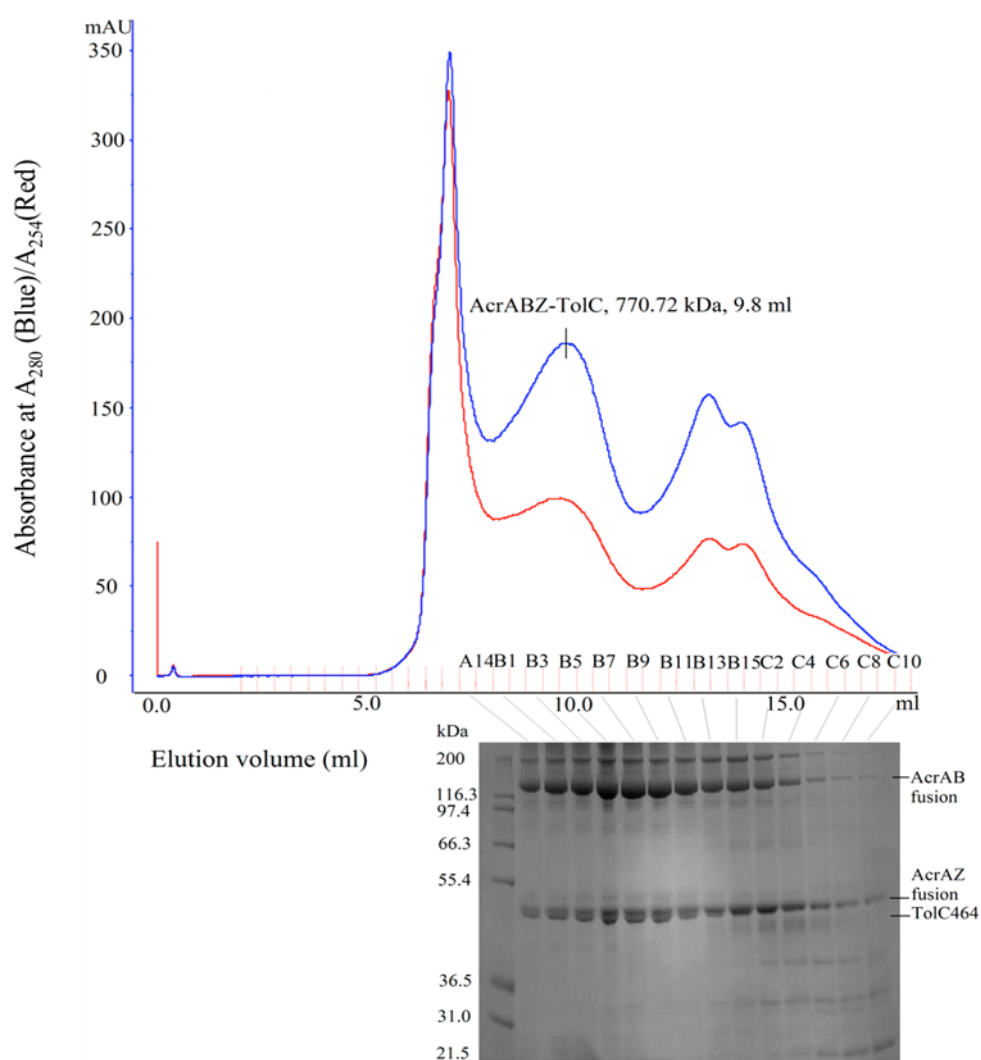
Extended Data Figure 2 | Validation of the AcrBZ crystal structure. An AcrB protomer taken from the refined structure of the AcrBZ–DARPin complex at a resolution of 3.3 Å was used in molecular replacement of the 3.7 Å data of the AcrBZ complex (Extended Data Table 1). The model of AcrB without AcrZ was refined with REFMAC³⁸ jelly-body and reference structure restraints, and a difference map calculated in which AcrZ does not contribute to the phases. Shown are the positive features of the difference map

in the vicinity of AcrZ (green carbon atoms) taken by superposing the AcrB protomers from the AcrBZ–DARPin and AcrBZ (red) structures. The unbiased map shows the presence of AcrZ and indicates that the presence of DARPin does not disrupt the interactions with AcrB. The model of the AcrBZ–DARPin complex at a resolution of 3.3 Å was refined without AcrZ to generate a simulated annealing omit map, which confirmed the location of AcrZ (data not shown). See also Fig. 1.



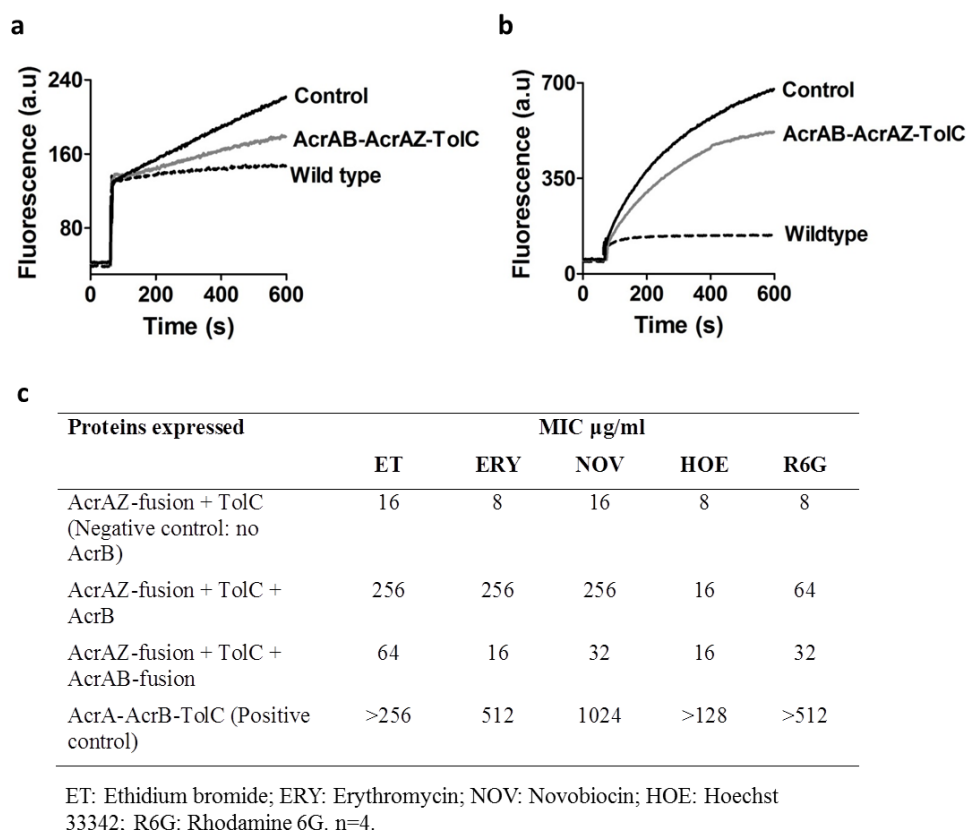
Extended Data Figure 3 | The AcrBZ complex resembles AcrB-YajC, and AcrBZ interactions involve conserved residues. **a**, Superposition of AcrB-YajC (orange and yellow; PDB accession 2RDD) onto AcrBZ (red and green). See also Fig. 1. We were not able to identify an interaction between AcrB and C-terminal histidine-tagged YajC from *E. coli*, suggesting that the interaction is less avid than the AcrB-AcrZ pairing (data not shown). It is interesting to note that the contacting surface is also conserved in another RND protein, SecDF, which is involved in protein translocation and membrane insertion⁴⁸. Therefore, it seems likely that SecDF might be modulated by a

similar helical peptide. Indeed, YajC forms a complex with SecDF^{49,50} and could act as an allosteric modulator in protein biogenesis. **b**, Sequence conservation of AcrZ. The secondary structure is annotated at the top and asterisks indicate residues that directly contact AcrB. **c**, Sequence variation of the surface of AcrB homologues, showing conservation of the surface that contacts AcrZ. Right-hand side includes the bound AcrZ in stick representation (green). The colour spectrum ranges from purple (most conserved) to blue (least conserved). This figure was made using ConSurf⁵¹.

a**b****c**

Extended Data Figure 4 | Schematic representations and purification of the fusion proteins used to assemble the efflux pump. **a**, The AcrA–AcrB fusion, with two flexible poly(Gly-Ser) linkers. The two C-terminal histidines of AcrB have been removed to prevent binding to the NTA matrix during co-purification. **b**, The AcrA–AcrZ–His₅ fusion and TolC co-expression construct. The numbers above the bars correspond to the residues of the

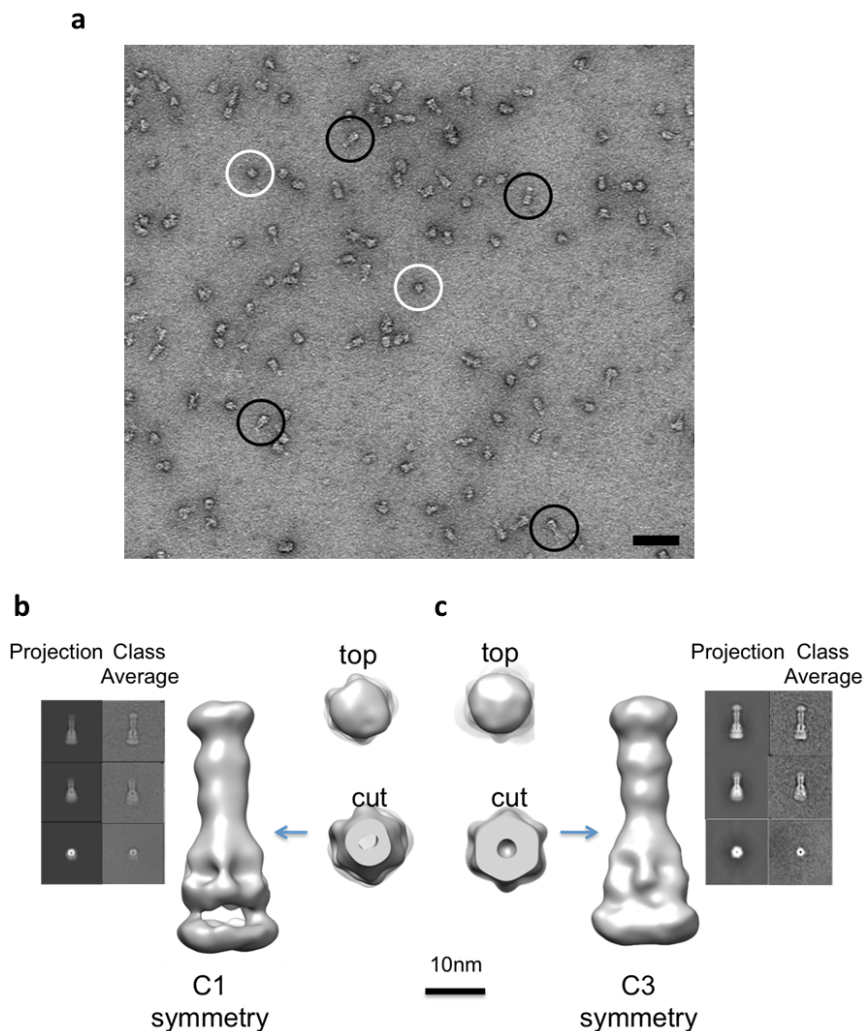
protein, and, owing to the restriction site used for cloning, a single glycine residue was inserted after the start codon in both AcrA and AcrZ. The flexible poly(Gly-Ser) linker permits the protomers to manoeuvre. **c**, Co-purification of TolC with the AcrABZ complexes. SDS–PAGE of the eluate from gel filtration following nickel affinity chromatography purification. See also Fig. 2.



Extended Data Figure 5 | The fusion assemblies can drive efflux *in vivo*.

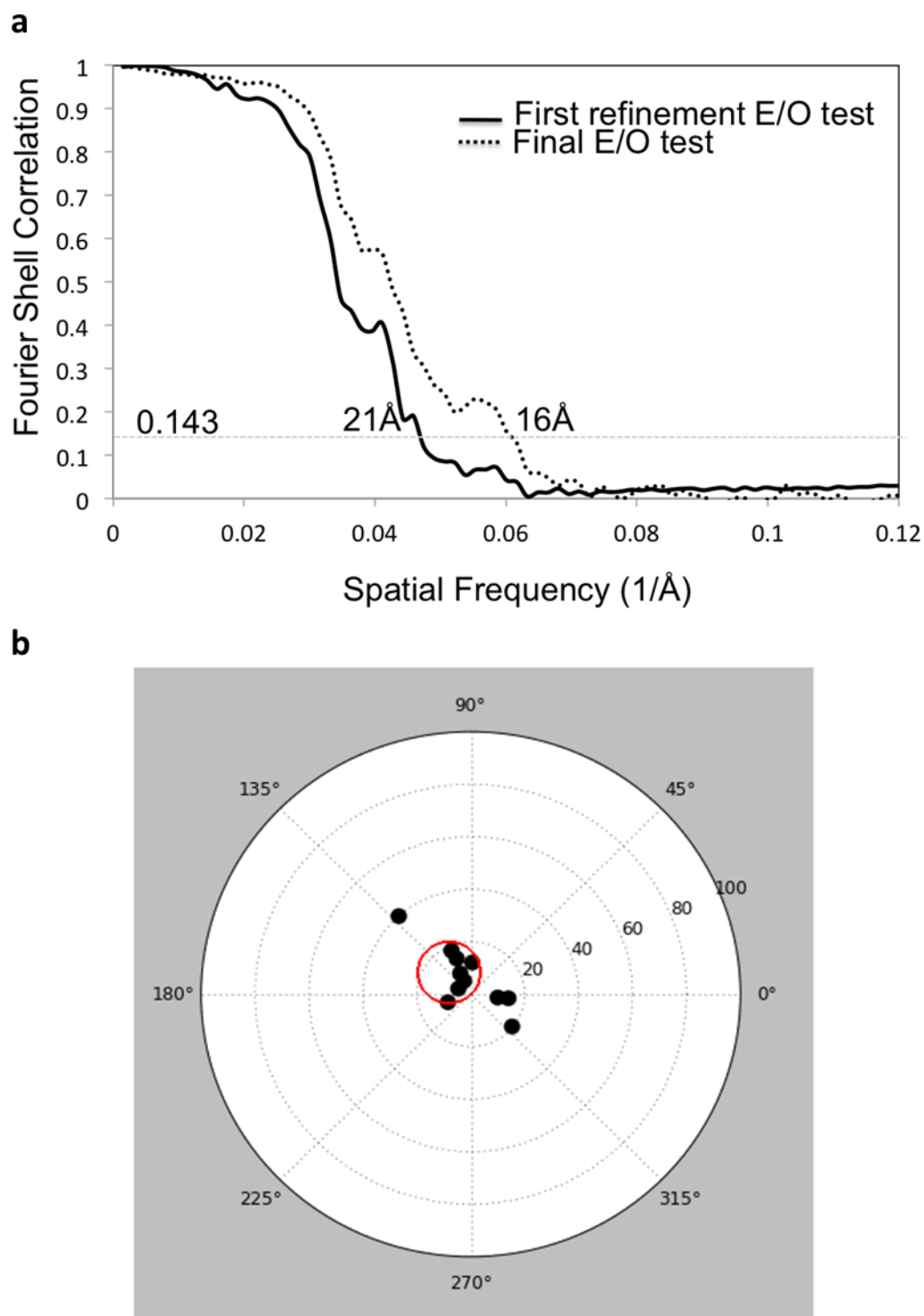
a, b, Drug-transport assays of the AcrAB and AcrAZ fusion proteins in the $\text{MC}\Delta\text{tolC}\Delta\text{acrAB}$ strain show that the engineered pump retains partial activity for efflux of ethidium bromide (**a**) and trimethylammonium-diphenylhexatriene (**b**). The results show representative traces from six biological replicates. For ethidium bromide, initial influx rates of 10.3 ± 0.6 , 6.4 ± 0.6 and 3.6 ± 0.4 arbitrary units (AU) per min were obtained for the non-expressing control cells, the cells expressing both fusion proteins plus TolC and the wild-type cells expressing AcrAB-TolC, respectively. For trimethylammonium-diphenylhexatriene, initial influx rates of 187 ± 6 , 151 ± 3 and 54 ± 2 AU min⁻¹ were obtained for the non-expressing control cells, the cells expressing both fusion proteins plus TolC and the wild-type cells

expressing AcrAB-TolC, respectively. **c**, MIC data on antimicrobial susceptibility of *E. coli* $\text{MC}\Delta\text{tolC}\Delta\text{acrAB}$ cells expressing wild-type AcrA-AcrB-TolC or the fusion proteins with TolC. These data indicate that the fusion of AcrA-AcrZ and the insertion of AcrA into AcrB both diminish the capacity for drug resistance. The plasmids pET21a, pET21a-AcrB or pET21a-*acrB*_{328-polyGS-*acrA*-polyGlySer-*acrB*}₃₂₉ (AcrAB), together with plasmid pRSFDuet-1-*acrA*-polyGlySer-*acrZ*_{H185}-tolC (AcrAZ + TolC), were transformed into $\text{MC}\Delta\text{tolC}\Delta\text{acrAB}$. Cells were tested for their ability to resist increasing concentrations of cytotoxic drugs. As a positive control, cells were transformed with plasmid pBAD_AcrA + AcrB + TolC, which encodes for the native AcrA-AcrB-TolC efflux pump.



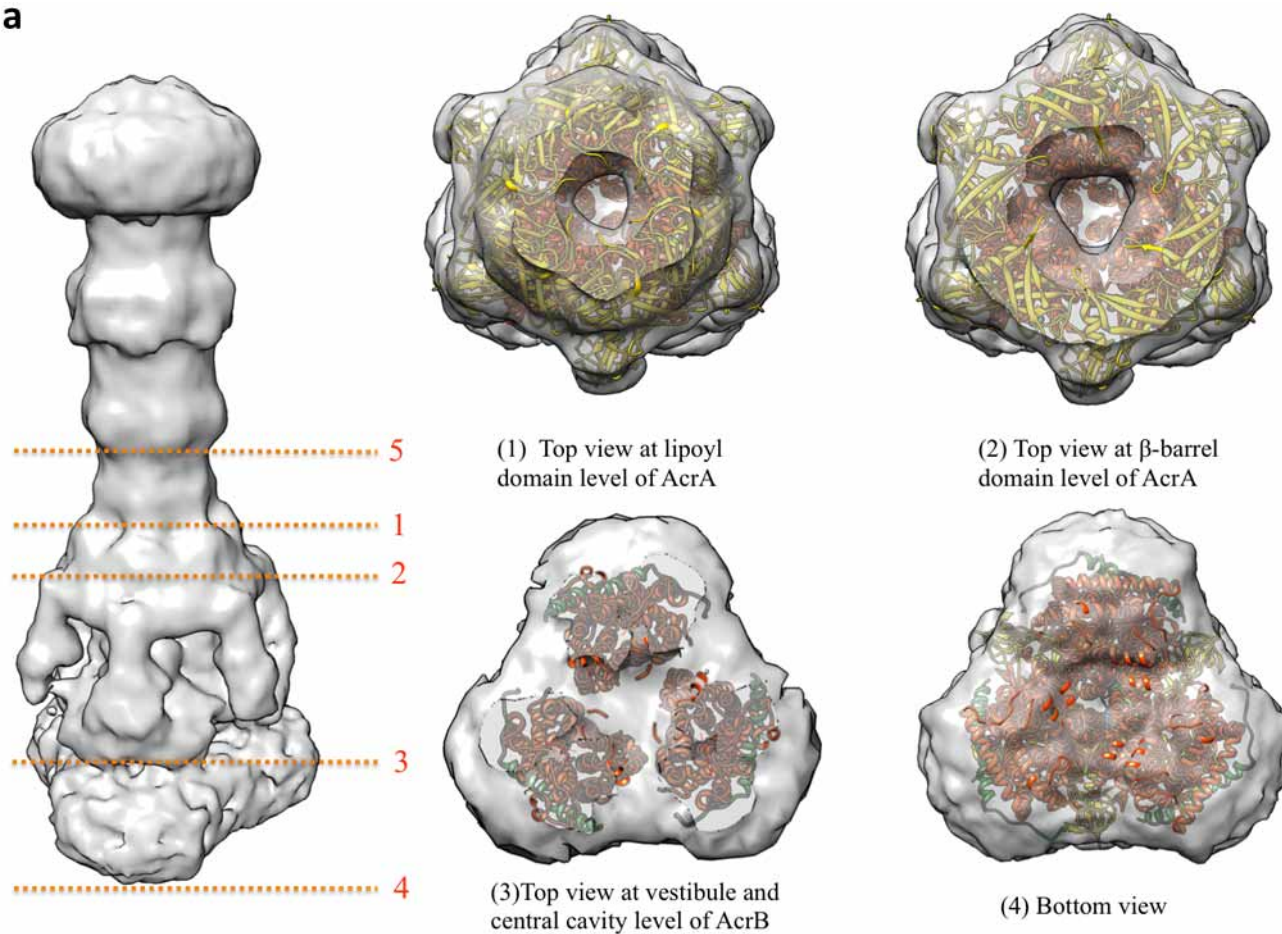
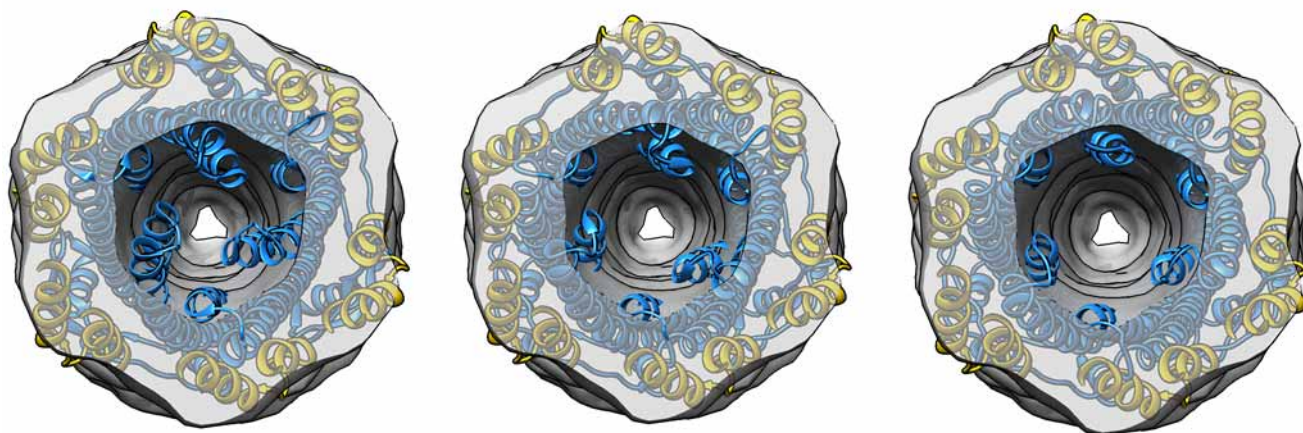
Extended Data Figure 6 | Electron microscopy images and class averages of the AcrABZ-TolC complex. **a**, Negative-stain electron microscopy images of the purified AcrABZ-TolC complex after GraFix treatment. White circles indicate particles with long axis almost normal to the viewing plane; black circles show particles with the long axis parallel to the viewing plane. **b**, **c**, Class averages from cryo-EM data and reconstructions with C1 symmetry and with C3 symmetry, respectively. The galleries in the side panels show representative two-dimensional class averages of the purified pump. The top

shows views perpendicular to the long axis of the drumstick shape, the middle shows inclined views, and the bottom shows views along the long axis. The reconstructed images and cross-sections indicate the presence of a six-fold symmetry, which is consistent with six AcrA protomers. The pseudo-atomic model has three protomers each of AcrB, AcrZ and TolC, and as TolC and AcrB each have an internal structural repeat, they have pseudo six-fold symmetry at low resolution (cut view). The maps are consistent with each other. See also Fig. 2.



Extended Data Figure 7 | Resolution estimation and validation of the cryo-EM map. **a**, Fourier shell correlation (FSC) of two independently determined cryo-EM maps of the efflux pump assembly after their alignment by Foldhunter⁵². **b**, Tilt-pair validation of the efflux pump assembly. Each point represents a pair of particles with experimentally known relative tilt.

The radius indicates the computationally determined amount of tilt, and the azimuth indicates tilt direction. The red circle denotes particle pairs that cluster around the experimental tilt axis geometry, thus validating the map in Fig. 2.

a**b**

Extended Data Figure 8 | Sections through the electron microscopy map.
a, Slices 1–4: view of planes normal to the three-fold symmetry axis of the pump. **b**, Slice 5 of the model assembly showing the opening of the TolC periplasmic domain. The same transverse plane through the cryo-EM map is

shown with TolC in the closed state, as seen in the crystal structure of the isolated protein (left; PDB accession 1EK9), the partially open structure, made by mutations (middle; PDB accession 2XMN) and the modelled fully opened structure (right).

Extended Data Table 1 | Crystallographic data and refinement statistics

Crystal	AcrB/AcrZ/DARPin	AcrB/AcrZ
Space group	R32 (H32)	R32 (H32)
Cell dimensions in hexagonal setting		
a, b, c (Å)	145.01, 145.01, 538.34	146.19, 146.19, 543.00
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	30.0-3.30 (3.46-3.30)*	24.89-3.69 (3.99-3.69)
R_{merge} , %	13.4 (98.6)	12.6 (117.0)
Total number of observations	166,656 (22,416)	244483 (50138)
Unique reflections	33,271 (4368)	23077 (4710)
Completeness, %	99.8 (100.0)	94.5 (94.8)
$I/\sigma I$ (outer shell)	11.0 (1.6)	15.1 (2.1)
$\langle I \rangle$ half-set correlation	0.998 (0.585)	0.999 (0.791)
B values Wilson plot/average in structure, Å ²	76.6/87.7	148/117.7
Reflections used in refinement	31,614	21,912
Geometry: Ramachandran favoured, allowed, outliers, %	95.2; 4.6; 0.2	94.9; 4.9; 0.3
Rms deviation from ideal values for bond lengths (Å) and bond angles (°)	0.012, 1.456	0.011, 1.57
Number of non-hydrogen atoms in refinement	9452	7997
Refinement (R/R_{free} 5% reflections in test set), %	28.3/32.4	33.9/36.1

*Values in parentheses are for highest-resolution shell

The PDB deposition codes for model and structure factors of AcrB/AcrZ/DARPin and AcrB/AcrZ are 4C48 and 4CDI respectively.

Structural basis of Sec-independent membrane protein insertion by YidC

Kaoru Kumazaki^{1,2*}, Shinobu Chiba^{3*}, Mizuki Takemoto^{1,2}, Arata Furukawa⁴, Ken-ichi Nishiyama⁵, Yasunori Sugano⁴, Takaharu Mori⁶, Naoshi Dohmae², Kunio Hirata⁷, Yoshiko Nakada-Nakura⁸, Andrés D. Maturana⁹, Yoshiki Tanaka⁴, Hiroyuki Mori¹⁰, Yuji Sugita⁶, Fumio Arisaka¹¹, Koreaki Ito³, Ryuichiro Ishitani^{1,2}, Tomoya Tsukazaki^{4,12} & Osamu Nureki^{1,2}

Newly synthesized membrane proteins must be accurately inserted into the membrane, folded and assembled for proper functioning. The protein YidC inserts its substrates into the membrane, thereby facilitating membrane protein assembly in bacteria; the homologous proteins Oxa1 and Alb3 have the same function in mitochondria and chloroplasts, respectively^{1,2}. In the bacterial cytoplasmic membrane, YidC functions as an independent insertase and a membrane chaperone in cooperation with the translocon SecYEG^{3–5}. Here we present the crystal structure of YidC from *Bacillus halodurans*, at 2.4 Å resolution. The structure reveals a novel fold, in which five conserved transmembrane helices form a positively charged hydrophilic groove that is open towards both the lipid bilayer and the cytoplasm but closed on the extracellular side. Structure-based *in vivo* analyses reveal that a conserved arginine residue in the groove is important for the insertion of membrane proteins by YidC. We propose an insertion mechanism for single-spanning membrane proteins, in which the hydrophilic environment generated by the groove recruits the extracellular regions of substrates into the low-dielectric environment of the membrane.

The Sec translocon, a protein-conducting channel conserved in all three phylogenetic domains, translocates secretory proteins across the membrane and inserts membrane proteins into the membrane by an hourglass-shaped pore formed by ten transmembrane helices^{6–8}. In bacteria, another membrane protein, YidC, is involved in the folding and insertion of many membrane proteins^{9–15}, such as subunit c of the F₀F₁-ATPase (F₀C), and thus is essential for cell viability^{3–5}. YidC is considered to function as both a Sec-dependent membrane chaperone and a Sec-independent insertase. In the Sec-independent pathway, YidC directly interacts with the translating ribosome to mediate the insertion of several single-spanning or double-spanning membrane proteins^{16,17}. Previous studies have suggested that YidC forms a face-to-face dimer, creating a channel at its dimer interface^{16,18}. By contrast, another research group has reported that a monomer of membrane-embedded YidC is sufficient for binding to the substrate-translating ribosome¹⁹. Oxa1, the mitochondrial homologue of YidC, reportedly functions as a voltage-gated membrane channel, as well as a membrane protein insertase, probably by forming a tetramer²⁰. However, the lack of a high-resolution structure of YidC has limited our understanding of the molecular mechanism of YidC-mediated membrane protein insertion.

Members of the genus *Bacillus* have two *yidC* genes, encoding YidC1 and YidC2. We determined the crystal structures of two constructs of *B. halodurans* YidC2 (YidC_{27–266} and YidC_{27–267}), which lack both the amino-terminal signal and the carboxy-terminal non-conserved sequences

(Fig. 1, Extended Data Table 1 and Extended Data Figs 1 and 2). A similar deletion variant of *Bacillus subtilis* YidC1 (also known as SpoIIIJ), an orthologue of *B. halodurans* YidC2 with 49.6% sequence identity, has comparable activity to the full-length protein *in vivo* (K248stop in Fig. 2a). Although the crystallographic asymmetric units of YidC_{27–266} and YidC_{27–267} contain one and two YidC molecules, respectively, the molecules in the crystalline lattice do not appear to form an effective oligomer, such as a face-to-face dimer^{16,18} (Extended Data Fig. 3a, b). Recent fluorescence correlation spectroscopy and cryo-electron microscopy analyses of YidC bound to a translating ribosome showed that YidC exists as a monomer in detergent solution and in lipid membranes^{17,19}. Consistent with these findings, our analysis using size exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) also showed that YidC exists as a monomer in detergent solution (Extended Data Fig. 3c–e). Because the overall structures of YidC_{27–266} and YidC_{27–267} are nearly identical (with a root mean square deviation of 1.79 Å over residues 27–266), we mainly describe the structure of YidC_{27–266}.

The YidC_{27–266} structure consists of the N-terminal E1 region, the C-terminal C3 tail and the core region: the core region is composed of transmembrane helices 1–5 (TM1–5), connected by two cytoplasmic (C1 and C2) regions and two extracellular (E2 and E3) regions (Fig. 1c). The E1 region consists of the EH1 helix, whereas the C1 region forms a hairpin-like structure composed of two helices (CH1 and CH2) connected by a short loop. The residues in the C2 region and the C3 tail are structurally disordered. The EH1, CH1 and CH2 helices protrude from the core region and lie nearly parallel to the plane of the membrane. The EH1 helix is amphipathic, whereas the CH1 and CH2 helices contain mainly hydrophilic residues. This observation suggested that one face of the EH1 helix is embedded in the membrane, whereas the CH1 and CH2 helices are exposed to the solvent (Fig. 1a). This idea is also consistent with theoretical calculations using an implicit membrane model (see Methods).

A comparison of the structures of YidC_{27–266} and YidC_{27–267} suggested that the C1 region, which does not interact with the remainder of YidC, is flexible (Extended Data Fig. 4a–c). The CH1 helix forms a continuous helix with the TM1 helix, which is kinked at the conserved proline residues P78 and P94 (Fig. 1c). The proline residues and the partially disordered flexible loop connecting the CH2 and TM2 helices (residues 130–140), which is likely to be embedded in the membrane, could enhance the flexibility of the C1 region. Indeed, the C1 region has higher B-factors than the other regions (Extended Data Fig. 5a). A molecular dynamics simulation also suggested that the position of the C1 region fluctuates greatly in the lipid bilayer environment (Extended Data Fig. 6a, b).

¹Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²Global Research Cluster, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan. ³Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan. ⁴Department of Systems Biology, Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan. ⁵Cryobiointerface Research Center, Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka, Iwate 020-8550, Japan. ⁶Theoretical Molecular Science Laboratory, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan. ⁷SR Life Science Instrumentation Unit, RIKEN Spring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan. ⁸Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ⁹Department of Bioengineering Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. ¹⁰Institute for Virus Research, Kyoto University, Shogoin Kawara-cho, Sakyo-ku, Kyoto 606-8507, Japan. ¹¹Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, Japan. ¹²JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.

*These authors contributed equally to this work.

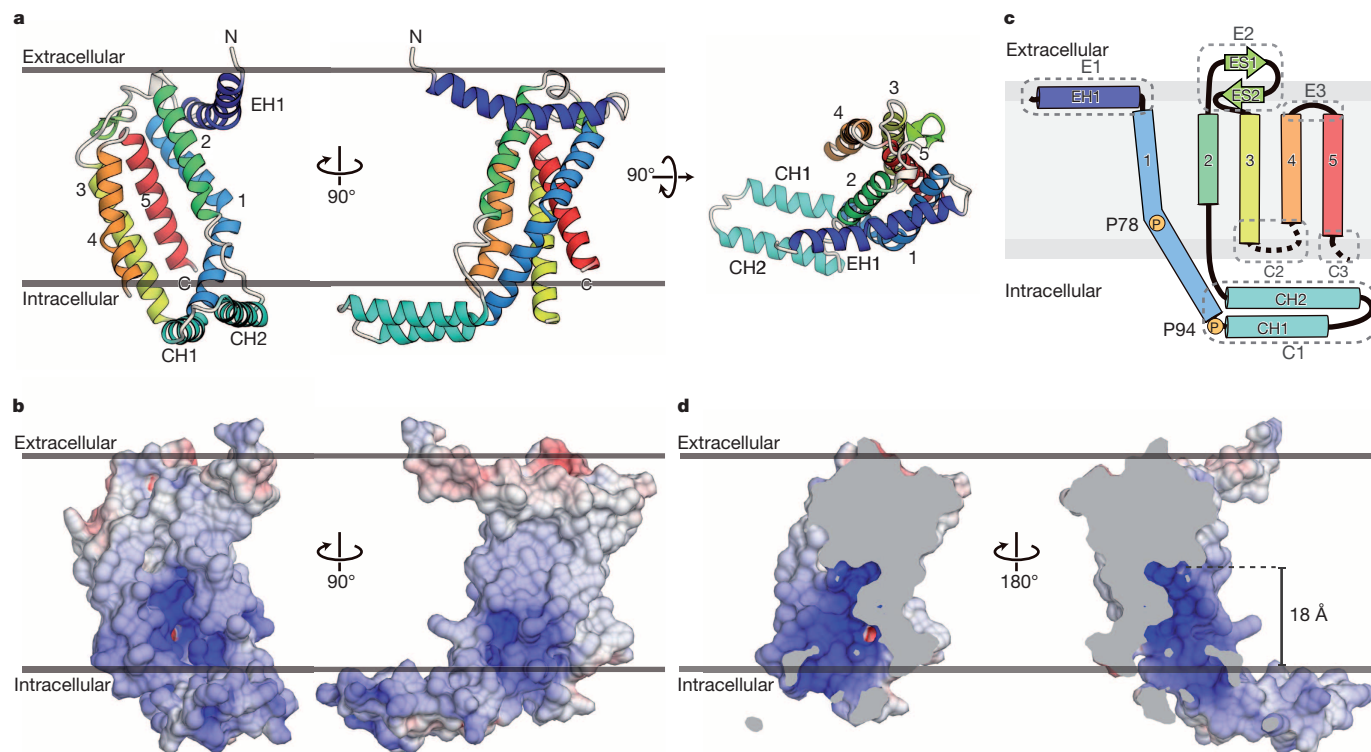


Figure 1 | Overall structure of YidC. **a**, Cartoon representations of *Bacillus halodurans* YidC viewed transversely through the membrane (left) and from the extracellular fluid (right). The structure is coloured from blue to red from the N to the C terminus. **b**, Surface model representations of YidC coloured according to electrostatic potential, ranging from blue (+20 kT/e) to red

(−20 kT/e), where kT is thermal energy and e is the elementary charge.

c, Topology diagram of YidC, coloured as in **a**. The grey area represents the cytoplasmic membrane. **d**, Cut-away molecular surface representations, viewed transversely through the membrane, coloured as in **b**.

To further investigate the functional importance of the C1 region, we performed a structure-based genetic analysis of *B. subtilis* SpoIIIJ. The membrane insertion of MifM, a single-spanning membrane protein, is mediated by the SpoIIIJ pathway²¹. In this analysis, we measured the β -galactosidase (LacZ) activity of a YidC2–LacZ translational fusion protein; this activity increased in response to a defect in the SpoIIIJ-dependent insertion of MifM (Extended Data Figs 7a and 8). Two C1-deletion mutants (Extended Data Fig. 7b) exhibited higher LacZ activities than wild-type SpoIIIJ (Fig. 2a), suggesting that the C1 region is crucial for the YidC-mediated membrane insertion of MifM. We then performed a growth complementation analysis using *B. subtilis* (as described in Extended Data Fig. 7c), which also suggested that the C1 region is important for SpoIIIJ activity (Fig. 2b).

The extracellular halves of the TM1–5 helices, in addition to the E2 region, are tightly packed with hydrophobic side chains (Extended Data Fig. 5b, c), whereas the cytoplasmic halves of the TM1–5 helices loosely interact with each other to form a groove ($\sim 2,000 \text{ \AA}^3$). This groove contains many hydrophilic residues, including the conserved T68, R72, Q82, Q142, Q187, N248 and Q254 residues, and thus generates a hydrophilic environment in the lipid bilayer (Fig. 2c and Extended Data Fig. 1). The conserved R72 residue, the only charged residue in this groove, protrudes into the centre of the groove and creates a strong positive electrostatic potential in the groove (Fig. 1b, d). This hydrophilic groove is open to both the cytoplasmic side and the membrane interior. By contrast, the groove is sealed towards the extracellular side by the hydrophobic core and is not accessible from that side (Fig. 1d). The hydrophobic core, which consists of the hydrophobic residues in the extracellular half of YidC, has lower B-factors than the other regions, suggesting that it is rigid (Extended Data Fig. 5). Structure-based genetic analyses suggested the importance of this hydrophobic core for the function of YidC (Fig. 2a–c and Extended Data Fig. 7b). A comparison of the YidC_{27–266} and YidC_{27–267} structures revealed that the groove in YidC_{27–267} is narrower

than that in YidC_{27–266}, suggesting that the groove has structural flexibility (Extended Data Fig. 4d). Molecular dynamics simulations also suggested that the size of the hydrophilic groove slightly fluctuated during a 1,000-ns simulation (Extended Data Fig. 6a). By contrast, the overall architecture of the core region, and the structure of the hydrophilic groove, remained stable in the lipid bilayer (Extended Data Fig. 6b), and the groove was constantly filled with ~ 20 water molecules (Extended Data Fig. 6c). The extracellular side of the groove remained sealed by the hydrophobic protein residues and the aliphatic lipid chains and thus was impermeable to ions and water molecules during molecular dynamics simulations. Taking these findings together, YidC appears to provide a flexible hydrophilic groove in the membrane that is open towards both the cytoplasmic side and the membrane interior but tightly sealed on the extracellular side.

To investigate the functional importance of the hydrophilic groove, we examined the membrane insertion activities of SpoIIIJ molecules in which R73 (R72 in YidC; Fig. 2c) or the conserved glutamine residues had been mutated, by using *in vivo* genetic analyses. All tested mutations of R73 abolished MifM insertion activity, except for replacement of arginine with lysine (R73K), which slightly decreased the insertion activity (Fig. 2d). None of the R73 mutants, except for R73K, complemented the growth of wild-type SpoIIIJ- and wild-type YidC2-depleted cells (Fig. 2b). By contrast, mutation of Q83, Q140, Q187 or Q238 (Q82, Q142, Q187 and Q254, respectively, in YidC; Fig. 2c) to alanine did not affect MifM insertion activity (Fig. 2e). A similar result was obtained for a chimera of the Pf3 coat protein, another single-spanning membrane protein inserted by the Sec-independent pathway, and the cytoplasmic region of MifM (Fig. 2f and Extended Data Fig. 9a). Taken together, these findings highlight the importance of the positive charge in the groove for insertion of MifM and the Pf3 coat protein into the membrane by YidC, whereas the conserved polar residues are probably important for creating the hydrophilic environment in the groove.

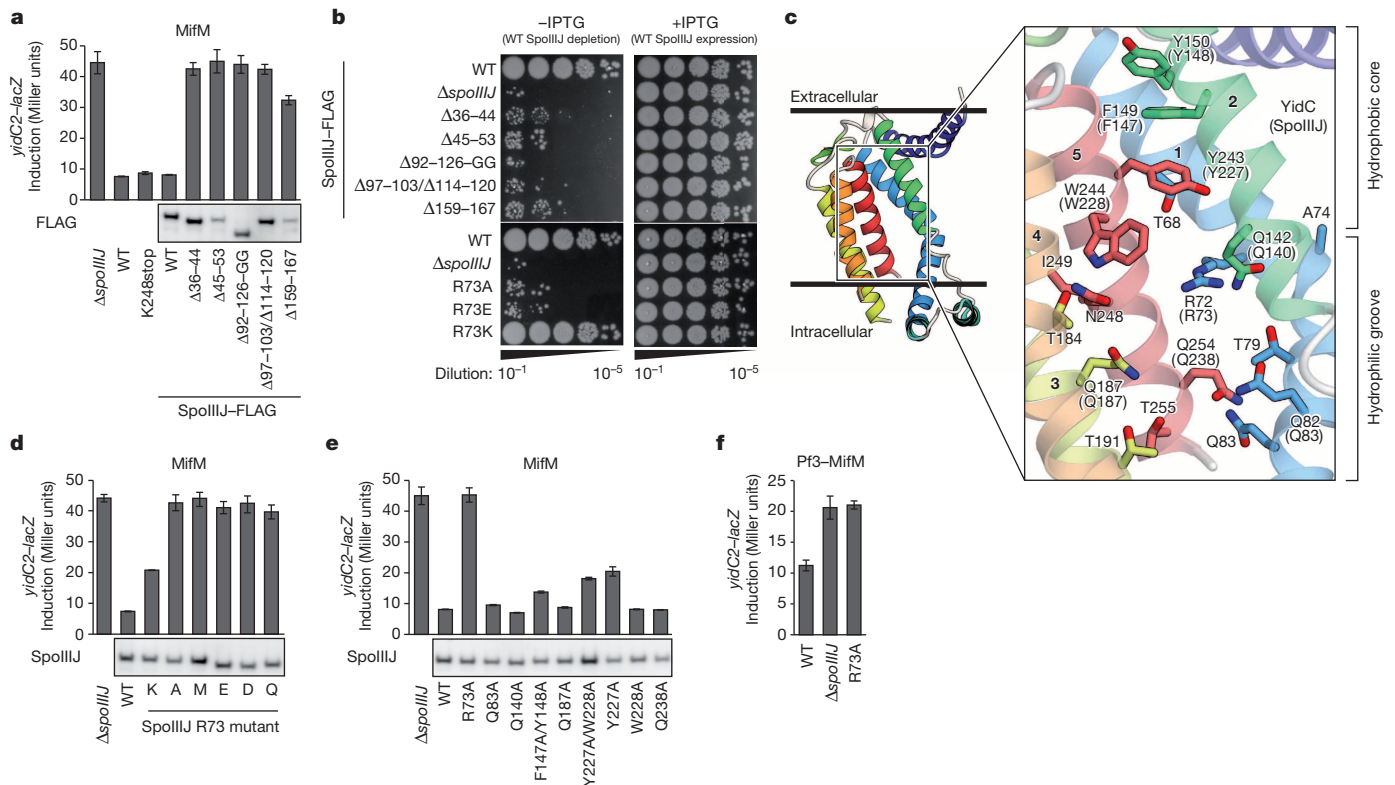


Figure 2 | In vivo functional analyses of YidC. **a**, **d**–**f**, Efficiencies of MifM and Pf3–MifM fusion protein insertion into the membrane, as determined by LacZ activity (mean \pm s.d., $n = 3$) in *spoIIIJ* mutant cells, as indicated (top). The accumulation of these proteins in the cell is shown, detected by anti-FLAG or anti-SpoIIIJ immunoblotting (bottom). K248stop, detected a *spoIIIJ* derivative that has a stop codon at position 248. Δ 92–126-GG represents a mutant in which the entire C1 region has been replaced by a glycine–glycine

linker. Δ 97–103/ Δ 114–120 represents a mutant in which both the CH1 and CH2 helices have been shortened by seven residues. **b**, Growth complementation of *Bacillus subtilis* cells reliant on a chromosomal SpoIIIJ mutant in the absence of isopropyl- β -D-thiogalactoside (IPTG). **c**, Close-up view of the hydrophilic groove of *Bacillus halodurans* YidC showing the side chains of the indicated residues. The corresponding residues in *B. subtilis* SpoIIIJ are indicated in parentheses. WT, wild type.

Several single-spanning membrane proteins, including MifM and the Pf3 coat protein, have acidic residues in their N-terminal extracellular tails. Thus, our results suggested that these acidic residues might interact with the arginine residue in the hydrophilic groove of YidC. To address this possibility, we examined the importance of these acidic residues by using an *in vivo* genetic analysis (Extended Data Fig. 9a). Mutating the acidic residues in the Pf3 coat protein had a less pronounced effect than mutating these residues in MifM; however, these mutations negatively affected the membrane insertion efficiencies of both MifM and the Pf3 coat protein, thereby supporting our hypothesis (Extended Data Fig. 9b). To further confirm that direct interactions occur between the substrate and the hydrophilic groove, we performed an *in vivo* site-directed ultraviolet-radiation crosslinking analysis, using *p*-benzoyl-L-phenylalanine (pBpa)²². pBpa was introduced at Q187 or W244 in the groove and at A74 or I249 on the exterior surface of YidC (Fig. 2c and Extended Data Fig. 1). *Escherichia coli* cells coexpressing MifM and the pBpa variant of YidC were exposed to ultraviolet radiation, and products that were crosslinked with MifM were analysed by immunoblotting. We detected crosslinked products only at positions 187 and 244, suggesting that the groove can interact with MifM (Fig. 3). Taken together, these observations suggested that the site around R72 in the hydrophilic groove is the substrate-binding site, through recognition of the hydrophilic residues of the substrate such as the acidic residues in MifM and the Pf3 coat protein.

On the basis of these structural and functional analyses, we propose a mechanism for the insertion of single-spanning membrane proteins with an acidic N-terminal extracellular region, such as MifM and the Pf3 coat protein, that is mediated by monomeric YidC (Fig. 4). In this mechanism, the substrate protein initially interacts with the C1 region (Fig. 4a) and is then transiently captured in the hydrophilic groove of YidC. This substrate binding can induce structural changes in the hydrophilic groove,

to accommodate the various substrate proteins. In this context, the conserved arginine residue in the groove might participate in substrate recognition (Fig. 4b). Subsequently, the transmembrane region of the substrate

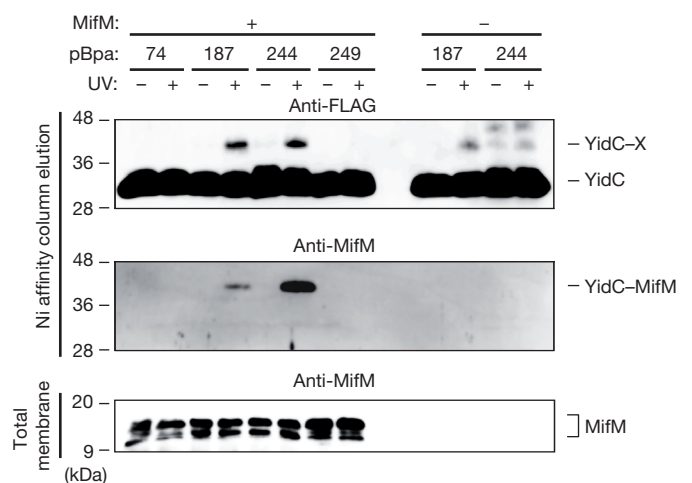


Figure 3 | Substrate binding to the hydrophilic groove. *In vivo* photo-crosslinking between the hydrophilic groove of YidC and MifM. The amino acid positions at which pBpa was introduced into YidC are indicated. Membrane proteins were purified by separation on a nickel affinity column and analysed by SDS-PAGE, and YidC-FLAG-His₆ and MifM were detected by anti-FLAG and anti-MifM immunoblotting, respectively. The ~40 kDa-bands are attributed to products in which YidC is crosslinked to MifM and to unidentified endogenous proteins, as indicated by YidC-X. The accumulation of MifM in the membrane is shown (bottom).

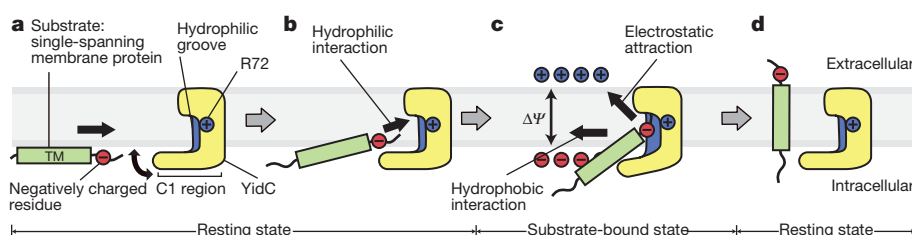


Figure 4 | Proposed model for membrane insertion of a single-spanning membrane protein. The obtained crystal structure probably represents the resting state before substrate binding. The hydrophilic region of the substrate would then be transiently captured in the hydrophilic groove of YidC (a, b), resulting in the substrate-bound state (c). Substrate release into the

membrane can be facilitated by the hydrophobic interaction between the transmembrane (TM) region and the lipid aliphatic chains, as well as by the membrane potential ($\Delta\psi$), attracting the negatively charged residues of the extracellular region of the single-spanning membrane protein by electrostatic attraction (c, d).

protein is released into the membrane, with the hydrophilic residues translocated to the extracellular side. The substrate release can be facilitated by the hydrophobic interaction between the transmembrane region and the lipid aliphatic chains. The membrane potential could also facilitate this process, by attracting the negative charge of the extracellular tail^{23,24} (Fig. 4c).

The above mechanism clearly explains the insertion of a certain class of membrane proteins: that is, single-spanning membrane proteins with an acidic N-terminal extracellular region. However, it cannot account for the YidC-mediated insertion of other classes of membrane protein. For example, for *E. coli* YidC, deletion of the C1 region or simultaneous substitution of five amino acids, including the conserved arginine residue, with serine, was not found to impair the insertion of an M13 procoat derivative, a double-spanning membrane protein²⁵. Therefore, it is likely that these membrane proteins are inserted by a different mechanism, which might involve either sites in the hydrophilic groove other than the arginine residue or transient oligomer formation by YidC. Further structural and biological studies are required to clarify the mechanism of the YidC-mediated insertion of these membrane protein classes.

METHODS SUMMARY

Histidine-tagged *Bacillus halodurans* YidC2 (YidC) was overproduced in *Escherichia coli*, solubilized from the membrane with *n*-dodecyl- β -D-maltoside and cholesteryl hemisuccinate and purified by successive nickel (Ni-NTA) chromatography and gel filtration chromatography steps. The histidine tag was then cleaved with tobacco etch virus (TEV) protease. Crystals were grown in a lipidic cubic phase, using monoolein. Diffraction data were collected on beamline BL32XU at SPring-8. The structure of YidC was determined by the multi-wavelength anomalous diffraction method, using a methyl-mercury-chloride-derivatized YidC (Y150C mutant) crystal and was refined to an $R_{\text{work}}/R_{\text{free}}$ of 24.2%/25.9% at 2.4 Å resolution. Molecular dynamics simulations of YidC in the explicit phosphoryl oleoyl phosphatidylethanolamine (POPE) lipid bilayer were performed with the NAMD program (version 2.8) for 1,000 ns. The MifM insertion activity of YidC was analysed by using a MifM-based assay in which the β -galactosidase (LacZ) activity of a YidC2-LacZ translational fusion increases in response to a defect in the SpoIIIJ-dependent insertion of MifM. The growth complementation assay was performed using mutant *Bacillus subtilis* cells, in which the *yidC2* gene on the chromosome was disrupted and the *spoIIIJ* gene was mutated, with rescue by a plasmid harbouring the wild-type isopropyl- β -D-thiogalactoside (IPTG)-inducible *spoIIIJ* gene. *In vivo* photo-crosslinking was performed using pBpa. *B. subtilis* MifM and pBpa variants of YidC were overexpressed in *E. coli*, and the cells were exposed to ultraviolet radiation. YidC was then purified by Ni-NTA chromatography, and products that were crosslinked with MifM were detected by immunoblotting. See Methods for further detail.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June 2013; accepted 24 February 2014.

Published online 16 April 2014.

1. Funes, S., Kauff, F., van der Sluis, E. O., Ott, M. & Herrmann, J. M. Evolution of YidC/Oxa1/Alb3 insertases: three independent gene duplications followed by functional specialization in bacteria, mitochondria and chloroplasts. *Biol. Chem.* **392**, 13–19 (2011).

2. Saller, M. J., Wu, Z. C., de Keyser, J. & Driessen, A. J. M. The YidC/Oxa1/Alb3 protein family: common principles and distinct features. *Biol. Chem.* **393**, 1279–1290 (2012).
3. Samuelson, J. C. et al. YidC mediates membrane protein insertion in bacteria. *Nature* **406**, 637–641 (2000).
4. Scotti, P. A. et al. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* **19**, 542–549 (2000).
5. Dalbey, R. E., Wang, P. & Kuhn, A. Assembly of bacterial inner membrane proteins. *Annu. Rev. Biochem.* **80**, 161–187 (2011).
6. Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu. Rev. Biophys.* **41**, 21–40 (2012).
7. van den Berg, B. et al. X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
8. Tsukazaki, T. et al. Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455**, 988–991 (2008).
9. Nagamori, S., Smirnova, I. N. & Kaback, H. R. Role of YidC in folding of polytopic membrane proteins. *J. Cell Biol.* **165**, 53–62 (2004).
10. Wagner, S. et al. Biogenesis of MalF and the MalFGK₂ maltose transport complex in *Escherichia coli* requires YidC. *J. Biol. Chem.* **283**, 17881–17890 (2008).
11. Yi, L. et al. YidC is strictly required for membrane insertion of subunits a and c of the F₁F₀ATP synthase and SecE of the SecYEG translocase. *Biochemistry* **42**, 10537–10544 (2003).
12. du Plessis, D. J. F., Nouwen, N. & Driessen, A. J. M. Subunit a of cytochrome *o* oxidase requires both YidC and SecYEG for membrane insertion. *J. Biol. Chem.* **281**, 12248–12252 (2006).
13. Price, C. E. & Driessen, A. J. M. Conserved negative charges in the transmembrane segments of subunit K of the NADH:ubiquinone oxidoreductase determine its dependence on YidC for membrane insertion. *J. Biol. Chem.* **285**, 3575–3581 (2010).
14. Serek, J. et al. *Escherichia coli* YidC is a membrane insertase for Sec-independent proteins. *EMBO J.* **23**, 294–301 (2004).
15. Facey, S. J., Neugebauer, S. A., Krauss, S. & Kuhn, A. The mechanosensitive channel protein MscL is targeted by the SRP to the novel YidC membrane insertion pathway of *Escherichia coli*. *J. Mol. Biol.* **365**, 995–1004 (2007).
16. Kohler, R. et al. YidC and Oxa1 form dimeric insertion pores on the translating ribosome. *Mol. Cell* **34**, 344–353 (2009).
17. Seitz, I., Wickles, S., Beckmann, R., Kuhn, A. & Kiefer, D. The C-terminal regions of YidC from *Rhodospirillum rubrum* and *Oceanicoccus anophageadensis* bind to ribosomes and partially substitute for SRP receptor function in *Escherichia coli*. *Mol. Microbiol.* **91**, 408–421 (2014).
18. Lotz, M., Haase, W., Kühlbrandt, W. & Collinson, I. Projection structure of yidC: a conserved mediator of membrane protein assembly. *J. Mol. Biol.* **375**, 901–907 (2008).
19. Kedrov, A. et al. Elucidating the native architecture of the YidC:ribosome complex. *J. Mol. Biol.* **425**, 4112–4124 (2013).
20. Krüger, V. et al. The mitochondrial oxidase assembly protein1 (Oxa1) insertase forms a membrane pore in lipid bilayers. *J. Biol. Chem.* **287**, 33314–33326 (2012).
21. Chiba, S., Lamsa, A. & Pogliano, K. A ribosome-nascent chain sensor of membrane protein biogenesis in *Bacillus subtilis*. *EMBO J.* **28**, 3461–3475 (2009).
22. Mori, H. & Ito, K. Different modes of SecY–SecA interactions revealed by site-directed *in vivo* photo-cross-linking. *Proc. Natl Acad. Sci. USA* **103**, 16159–16164 (2006).
23. Chen, M. et al. Direct interaction of YidC with the Sec-independent PF3 coat protein during its membrane protein insertion. *J. Biol. Chem.* **277**, 7670–7675 (2002).
24. Zhu, L. L., Wasey, A., White, S. H. & Dalbey, R. E. Charge-composition features of model single-span membrane proteins that determine selection of YidC and SecYEG translocase pathways in *Escherichia coli*. *J. Biol. Chem.* **288**, 7704–7716 (2013).
25. Jiang, F. et al. Defining the regions of *Escherichia coli* YidC that contribute to activity. *J. Biol. Chem.* **278**, 48965–48972 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We wish to thank K. Watanabe from Shoko Scientific for assistance with the SEC-MALLS experiments; T. Nishizawa, T. Higuchi, H. E. Kato, M. Hattori, R. Ishii and H. Nishimura for discussions; A. Kurabayashi, H. Nakamura, S. Hibino, T. Takino and C. Tsutsumi for technical support; A. Nakashima and R. Yamazaki for secretarial assistance; the RIKEN BioResource Center for providing

B. halodurans genomic DNA; the RIKEN Integrated Cluster of Clusters (RICC) for providing computational resources; and the beamline staff members at BL32XU of SPring-8 for technical assistance during data collection. The synchrotron radiation experiments were performed at BL32XU of SPring-8 (proposal no. 2011A1125, 2011A1139, 2011B1062, 2011B1280, 2012A1093, 2012A1201, 2012B1146, 2012B1162 and 2013A1128), with approval from RIKEN. This work was supported by the Platform for Drug Discovery, Informatics and Structural Life Science by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), by JSPS KAKENHI (grant no. 20247020, 20523517, 24687016, 24102503, 24121704, 24227004, 24657095, 25291006, 25291009 and 25660073), by the FIRST program, by PRESTO, by the JST, by a Grant-in-Aid for JSPS Fellows, by a grant for the HPCI STRATEGIC PROGRAM Computational Life Science and Application in Drug Discovery and Medical Development from MEXT, and by grants from the Private University Strategic Research Foundation Support Program (MEXT), the Nagase Science and Technology Foundation, and the Astellas Foundation for Research on Metabolic Disorders.

Author Contributions K.K. performed the crystallization and structure determination. S.C. performed the genetic analyses. K.K., A.F., K.-I.N., Y. Sugano, A.D.M., Y.T., H.M. and T.T. performed the functional analysis. M.T., T.M., Y. Sugita and R.I. performed the molecular dynamics simulation. K.K., N.D. and F.A. identified the molecular mass. K.H., Y.N.-N., R.I., T.T. and O.N. assisted with the structure determination. K.K., S.C., K.I., R.I., T.T. and O.N. wrote the manuscript. T.T. and O.N. directed and supervised all of the research.

Author Information The atomic coordinates and structure factors for YidC_{27–266} and YidC_{27–267} have been deposited in the Protein Data Bank under accession numbers 3WO6 and 3WO7, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. (nureki@bs.s.u-tokyo.ac.jp) or T.T. (ttsukaza@bs.naist.jp).

METHODS

Cloning, expression and purification of YidC for structure determination. *Bacillus halodurans* YidC2 (YidC) was cloned into a pET-modified vector²⁶. The two resultant plasmids, encoding YidC_{1–26}-H₈-ENLYFQGG-YidC_{27–266} (YidC_{27–266}) and YidC_{1–26}-ENLYFQGG-YidC_{27–267}-LESSV-ENLYFQGG-green fluorescent protein (GFP)-H₈ (YidC_{27–267}), were expressed in *Escherichia coli* C41 (DE3) cells harbouring pRARE (Novagen), and the proteins were purified using the following protocol. The cells were grown in a 5 l culture at 37 °C to an absorbance at 600 nm of 0.7, and expression was induced with 1 mM isopropyl- β -D-thiogalactoside (IPTG) at 15 °C for 16 h. The cells were then harvested by centrifugation (4,500 g, 10 min, 4 °C), resuspended in buffer containing 20 mM Tris-HCl, pH 8.0, and 0.1 mM phenylmethylsulphonyl fluoride, and disrupted by two passages through a Microfluidizer processor (Microfluidics) at 15,000 p.s.i. After removal of the debris by centrifugation (12,000 g, 30 min, 4 °C), the supernatant was ultracentrifuged (138,000 g, 1 h, 4 °C) to pellet the membranes, which were then solubilized in a buffer containing 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 20 mM imidazole, 1% *n*-dodecyl- β -D-maltoside (DDM) and 0.1% cholesteryl hemisuccinate (CHS). Insoluble materials were removed by ultracentrifugation (138,000 g, 30 min, 4 °C), and the supernatant was mixed with 5 ml Ni-NTA Superflow (Qiagen). After binding for 1 h, the resin was washed with 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 20 mM imidazole, 0.1% DDM and 0.01% CHS, and YidC was eluted with the same buffer supplemented with imidazole to a final concentration of 300 mM at 4 °C. The N-terminal residues and the His₈ tag or GFP-His₈ tag were cleaved by a His-tagged tobacco etch virus (TEV) protease, and the sample was reloaded onto the Ni-NTA column (5 ml) to remove the protease. The flow-through fraction containing YidC was collected, concentrated and further purified on a Superdex 200 10/300 GL column (GE Healthcare) in 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 0.1% DDM and 0.01% CHS at 4 °C. For crystallization, the purified protein was concentrated to 6 mg ml⁻¹ with a centrifugal filter device (Millipore, 50 kDa molecular weight cutoff) and dialysed against the following buffer (1 mM Tris-HCl, pH 8.0, 0.05% DDM and 0.005% CHS). For mercury derivatization, the Y150C YidC_{27–266} mutant was purified and incubated with 2 mM methyl mercury chloride for 1 h before crystallization.

Crystallization and heavy-atom derivatization. The protein was mixed with monoocten in a 2:3 protein to lipid ratio (w/w), using the twin-syringe mixing method²⁷. Aliquots (50 nl) of the protein lipidic cubic phase (LCP) mixture were spotted on a 96-well sandwich plate and overlaid with 800 nl precipitant solution using a mosquito LCP crystallization robot (TTP LabTech). The crystals of YidC_{27–266} and YidC_{27–267} were grown at 20 °C in reservoir solutions containing 28–32% poly(ethylene glycol) 500 dimethylether (PEG500DME), 2.5 mM CdCl₂ and 100 mM Na(CH₃)₂AsO₂, pH 6.0; and 24–26% PEG500DME, 10 mM CuCl₂, 200 mM NH₄COOH and 100 mM MES-NaOH, pH 6.0, respectively. The heavy-atom derivatized crystals were obtained by co-crystallization of the Y150C mutant and methyl mercury chloride in the same reservoir solution used for the YidC_{27–266} crystals. The crystals grew to full size in 2–3 weeks. The crystals were flash-cooled, using reservoir solution supplemented with 20% PEG500DME and 20% glycerol as a cryoprotectant and then stored in liquid nitrogen.

Data collection and structure determination. X-ray diffraction data sets were collected by the helical data collection method on beamline BL32XU at SPRING-8, using a micro beam with a 1- μ m width and a 10- μ m height²⁸. Diffraction data were processed using the program HKL-2000 (HKL Research) or the program XDS²⁹. One Hg atom site was identified with the program SHELXD³⁰. The initial phases were calculated using the program SHARP³¹, followed by solvent flattening with SOLOMON³². The main chain was traced by automated model building using the program RESOLVE³³. The model was further built manually using COOT³⁴ and refined using PHENIX³⁵. The structures of YidC_{27–266} and YidC_{27–267} were determined by molecular replacement, using the program PHASER³⁶. The Ramachandran plots were calculated with the program RAMPAGE³⁷. Data collection and refinement statistics are provided in Extended Data Table 1. The YidC_{27–266} crystal contains one molecule in the asymmetric unit. The YidC_{27–267} crystal contains two molecules in the asymmetric unit (Mol A and Mol B), but these molecules do not form a face-to-face dimer. The figures depicting the molecular structures were prepared using CueMol (<http://www.cuemol.org/>).

Bacterial strains and plasmids for *in vivo* functional analysis. The *B. subtilis* strains, plasmids and DNA oligonucleotides used are listed in Supplementary Tables 1–3 and were constructed as described in the Supplementary Methods.

β -Galactosidase activity assay and immunoblotting. *B. subtilis* cells were cultured at 37 °C in LB or CH medium. Aliquots (500 μ l) of cultures at an optical density at 600 nm of ~0.5 were harvested and used for β -galactosidase activity assays and immunoblotting. The β -galactosidase activities were measured as described previously^{38,39}. For immunoblotting, a 500 μ l aliquot of the culture was mixed with 56 μ l 50% trichloroacetic acid and incubated on ice for at least 5 min. The cells were precipitated by centrifugation (4 °C, 15,000 r.p.m., 5 min), washed with 1 ml 1 M Tris-HCl, pH 8.0, resuspended in SB buffer (33 mM Tris-HCl, pH 8.0, 40% sucrose and 1 mM EDTA)

containing 1 mg ml⁻¹ lysozyme (Sigma) and incubated at 37 °C for 10 min. The cells were then solubilized by adding an equal volume of 2 \times SDS loading buffer (4% SDS, 0.1 M Tris-HCl, pH 6.8, 30% glycerol and 10 mM dithiothreitol (DTT)) and subjected to immunoblotting, using either anti-FLAG (Sigma) or anti-SpoIII antibodies, as described previously⁴⁰. Antiserum production is described in the Supplementary Information.

Growth complementation assay. *B. subtilis* cells were cultured at 37 °C in LB medium containing 100 μ g ml⁻¹ spectinomycin and 1 mM IPTG. Aliquots (3.5 μ l) of serially diluted (10⁻¹–10⁻⁵), fully grown cultures were spotted on spectinomycin-containing LB agar plates with or without 1 mM IPTG and incubated at 37 °C for 15 h.

***In vivo* photo-crosslinking assay.** *B. halodurans* yidC-FLAG (Sigma-Aldrich)-His₈ and *B. subtilis* mifM were cloned into the NcoI and BamHI sites of MCS1 and the NdeI and XhoI sites of MCS2 in pETDuet (Novagen), respectively. An amber mutation, TAG, was introduced into yidC by site-directed mutagenesis. mifM-deletion plasmids were generated by restriction digestion of the plasmids with SalI and XhoI and then ligation. *E. coli* BL21 (DE3) cells harbouring two plasmids, pEVOL-pBpF (Addgene) and a pETDuet-based plasmid expressing YidC and MifM, were grown at 37 °C in M9-glucose medium supplemented with 1 mM pBpa²² and appropriate antibiotics until mid-logarithmic phase, and then induced with 1 mM IPTG for 30 min. Portions of the culture (1 ml for isolation of total membrane and 2 ml for purification of YidC by Ni-NTA chromatography) were transferred to a dish and exposed to ultraviolet radiation (365 nm) for 5 min, by using a B-100AP ultraviolet lamp (UVP) at a distance of 5 cm. The irradiated cells were collected by centrifugation at 7,000 g for 1 min at 4 °C, suspended in 300 μ l 10 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA-Na and 0.1 mM 4-(2-aminoethyl)benzenesulphonyl fluoride hydrochloride, and disrupted by freeze-thawing and sonication (Qsonica) with cooling on ice. Cellular debris was separated by centrifugation at 9,000 g for 1 min at 4 °C and used for the isolation of the total membrane fraction or for the purification of YidC. The total membrane fraction was isolated by ultracentrifugation at 100,000 g for 20 min at 4 °C. For the purification of YidC, the membrane proteins were solubilized in SC buffer containing 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole-HCl, pH 8.0, 1% DDM, 0.1% CHS and 0.1 mM 4-(2-aminoethyl)benzenesulphonyl fluoride hydrochloride for 30 min at 4 °C. Insoluble materials were removed by ultracentrifugation at 100,000 g for 20 min at 4 °C. The supernatant was mixed with 0.1 ml Ni-NTA agarose (Qiagen) that had been pre-equilibrated with SC buffer. After binding for 30 min at 4 °C, the resin was washed with SC buffer containing 0.1% DDM, and then YidC was eluted with SC buffer supplemented with 300 mM imidazole-HCl (pH 8.0). Proteins were separated by SDS-PAGE and detected by immunoblotting using anti-FLAG (Sigma) or anti-MifM⁴⁰ antibodies.

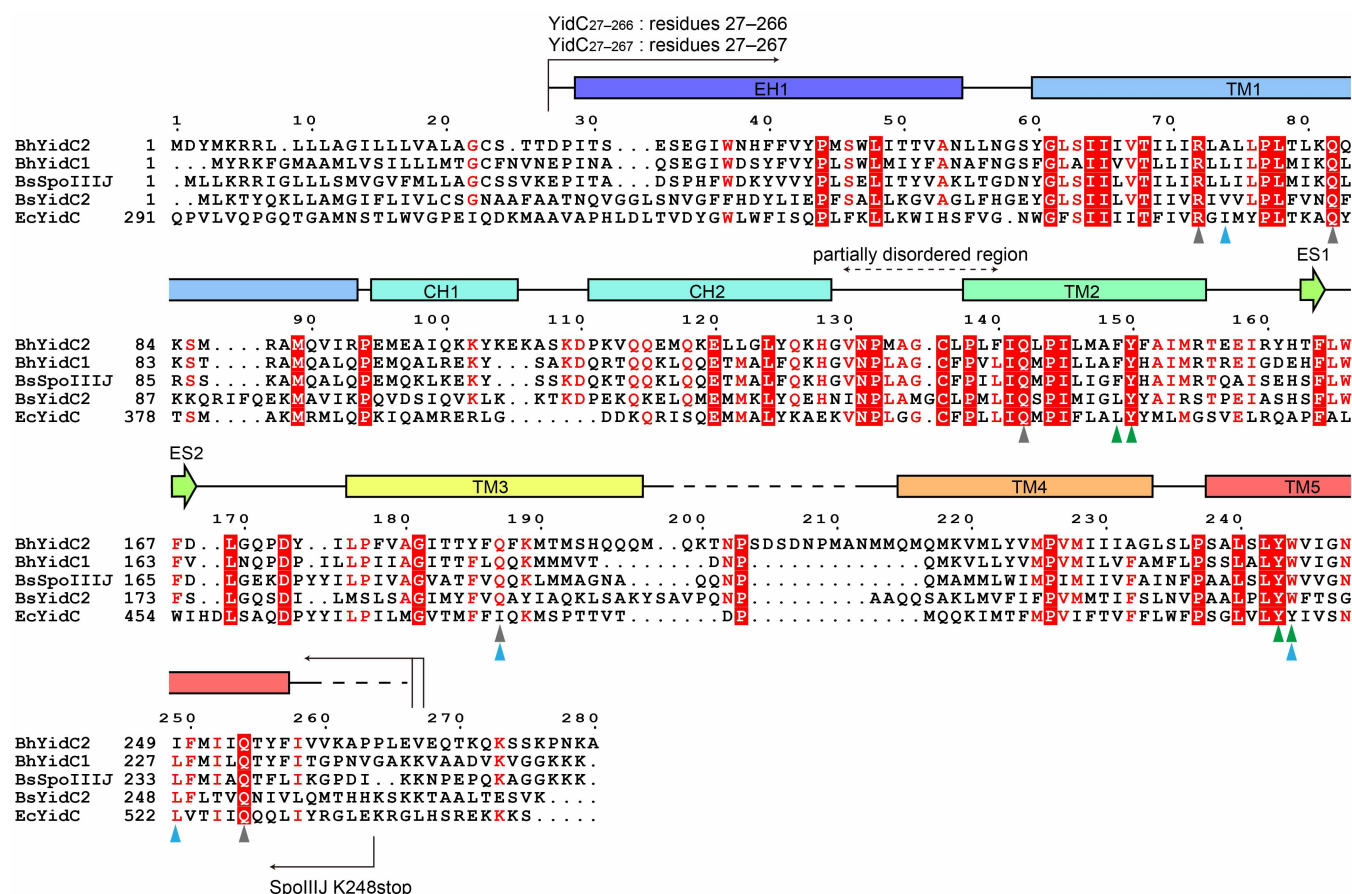
Molecular dynamics simulations. The simulation system included YidC, phosphoryl oleoyl phosphatidylethanolamine (POPE), water molecules and 150 mM NaCl. At first, the position and orientation of YidC in the POPE lipid bilayer were optimized using implicit solvent and membrane models (T.M. and Y. Sugita, manuscript in preparation). Next, the disordered region in the C2 loop (residues 200–213) was modelled, using the program MODELLER⁴¹. The missing atoms, including hydrogens in the protein, were built with the program VMD⁴². Finally, the periodic boundary system, including the explicit solvent and the POPE lipid bilayer⁴³, was prepared. The resultant size of the simulation box was 96 Å \times 96 Å \times 96 Å. The net charge of the solute was neutralized by adding chloride and sodium ions. The molecular topologies and parameters from CHARMM36 were used⁴³.

Molecular dynamics simulations were performed with the program NAMD (version 2.8)⁴⁴. The system was first energy minimized for 1,000 steps with fixed positions for the non-hydrogen atoms, and then for another 1,000 steps with 10 kcal mol⁻¹ restraints for the non-hydrogen atoms. Next, we performed a long equilibration run of 50 ns in the canonical (NVT) ensemble (300 K, 96 Å \times 96 Å \times 96 Å volume), with 10 kcal mol⁻¹ restraints for protein non-hydrogen atoms and 0.1 kcal mol⁻¹ restraints for water molecules, to optimize the locations of the lipid molecules around the protein, especially around the hydrophilic groove. Finally, equilibration was performed for 5 ns in the isothermal-isobaric (NPT) ensemble (300 K, 1.0 atm), with 10 and 0.1 kcal mol⁻¹ restraints for the protein main chain and side chain atoms, respectively. The production process was performed for 1,000 ns. Constant temperature was maintained by using Langevin dynamics. Constant pressure was maintained by using the Langevin piston Nosé-Hoover method⁴⁵. Long-range electrostatic interactions were calculated using the particle mesh Ewald method⁴⁶.

Determination of the molecular mass of YidC in detergent solution by using SEC-MALLS. The instrument set-up used for the SEC-MALLS experiment consisted of a Prominence HPLC system (Shimadzu) with an SPD-20A ultraviolet absorbance detector connected in series with a DAWN HELEOS II light-scattering detector (Wyatt Technology) and an Optilab T-REX refractive index detector (Wyatt Technology). Analytical size exclusion chromatography was performed using a Superdex 200 10/300 column (GE Healthcare) equilibrated with buffer containing 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 0.1% DDM and 0.01% CHS. Purified YidC_{27–266}

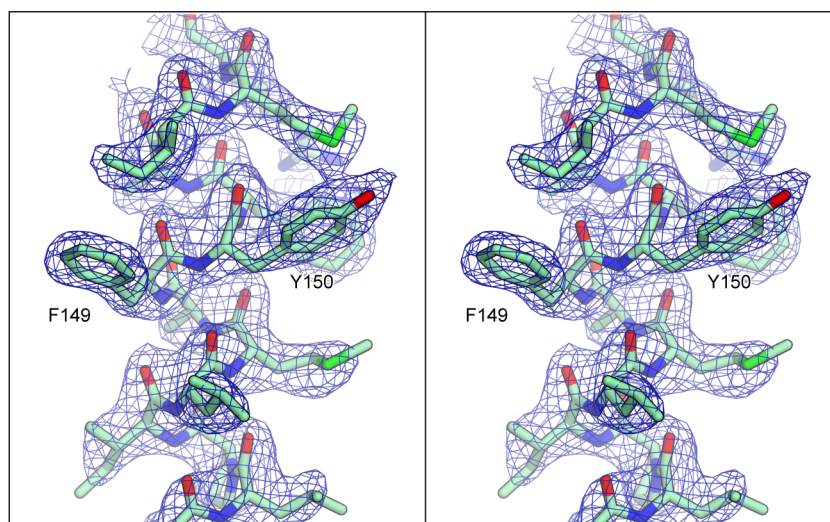
(60 µg) was injected onto the column, and the elution was monitored in-line with three detectors. A 658.0-nm wavelength laser was used in the light scattering experiment. The data were corrected for the volume delay of ultraviolet radiation between the other detectors and were analysed using ASTRA software (Wyatt Technology). The molecular masses of the protein–micelle complex, the micelle and the protein were determined as described previously^{47,48}. The refractive index increment (dn/dc) in which n is the refractive index and c is the concentration of the mixture of DDM and CHS in buffer, containing 300 mM NaCl and 20 mM Tris-HCl, pH 8.0, was determined offline using an Optilab T-rEX refractive index detector with a 658.0-nm wavelength laser, as described previously⁴⁹.

26. Nishizawa, T. *et al.* Structural basis for the counter-transport mechanism of a H^+/Ca^{2+} exchanger. *Science* **341**, 168–172 (2013).
27. Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009).
28. Hirata, K. *et al.* Achievement of protein micro-crystallography at SPring-8 beamline BL32XU. *J. Phys. Conf. Ser.* **425**, 012002 (2013).
29. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
30. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
31. de La Fortelle, E., Irwin, J. J. & Bricogne, G. SHARP: a maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
32. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
33. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
37. Lovell, S. C. *et al.* Structure validation by $C\alpha$ geometry: ϕ , ψ and $C\beta$ deviation. *Proteins* **50**, 437–450 (2003).
38. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, 1972).
39. Rubio, A., Jiang, X. & Pogliano, K. Localization of translocation complex components in *Bacillus subtilis*: enrichment of the signal recognition particle receptor at early sporulation septa. *J. Bacteriol.* **187**, 5000–5002 (2005).
40. Chiba, S. *et al.* Recruitment of a species-specific translational arrest module to monitor different cellular processes. *Proc. Natl Acad. Sci. USA* **108**, 6073–6078 (2011).
41. Sali, A. *et al.* Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318–326 (1995).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
43. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
44. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
45. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613 (1995).
46. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
47. Hayashi, Y., Matsui, H. & Takagi, T. Membrane protein molecular weight determined by low-angle laser light-scattering photometry coupled with high-performance gel chromatography. *Methods Enzymol.* **172**, 514–528 (1989).
48. Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73–82 (2008).
49. Strop, P. & Brunger, A. T. Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci.* **14**, 2207–2211 (2005).
50. Chiba, S. & Ito, K. Multisite ribosomal stalling: a unique mode of regulatory nascent chain action revealed for MifM. *Mol. Cell* **47**, 863–872 (2012).



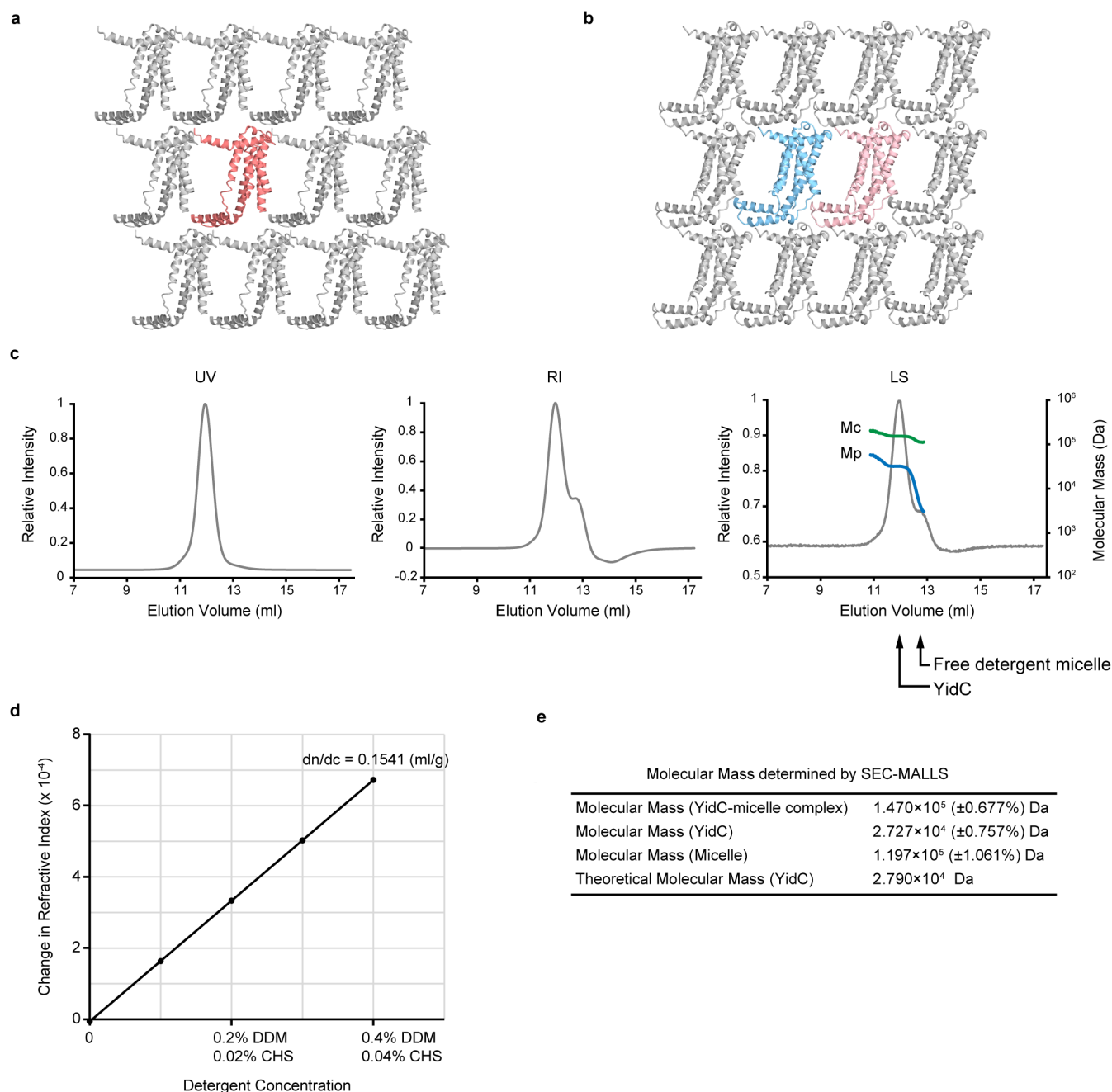
Extended Data Figure 1 | Multiple amino acid sequence alignment of YidC proteins. Sequence alignment of *Bacillus halodurans* YidC2 (BhYidC2), *B. halodurans* YidC1 (BhYidC1), *Bacillus subtilis* SpoIIIJ (BsSpoIIIJ), *B. subtilis* YidC2 (BsYidC2) and *Escherichia coli* YidC (EcYidC). The secondary structure of YidC₂₇₋₂₆₆ is indicated above the sequences. The α -helices (as described in the main text) and β -strands (ES1 and ES2 in the E2 region) are indicated by cylinders and arrows, respectively. Strictly conserved residues among the five

molecules are highlighted in red boxes, and highly conserved residues are indicated by red letters. The hydrophilic and bulky residues that were mutated and the pBpa positions introduced into *B. halodurans* YidC2 are indicated by grey, green and blue triangles, respectively. The *spoIIIJ* K248stop derivative has a stop codon introduced at position 248, as indicated, and thereby expresses a SpoIIIJ mutant that lacks the C-terminal 14 residues.



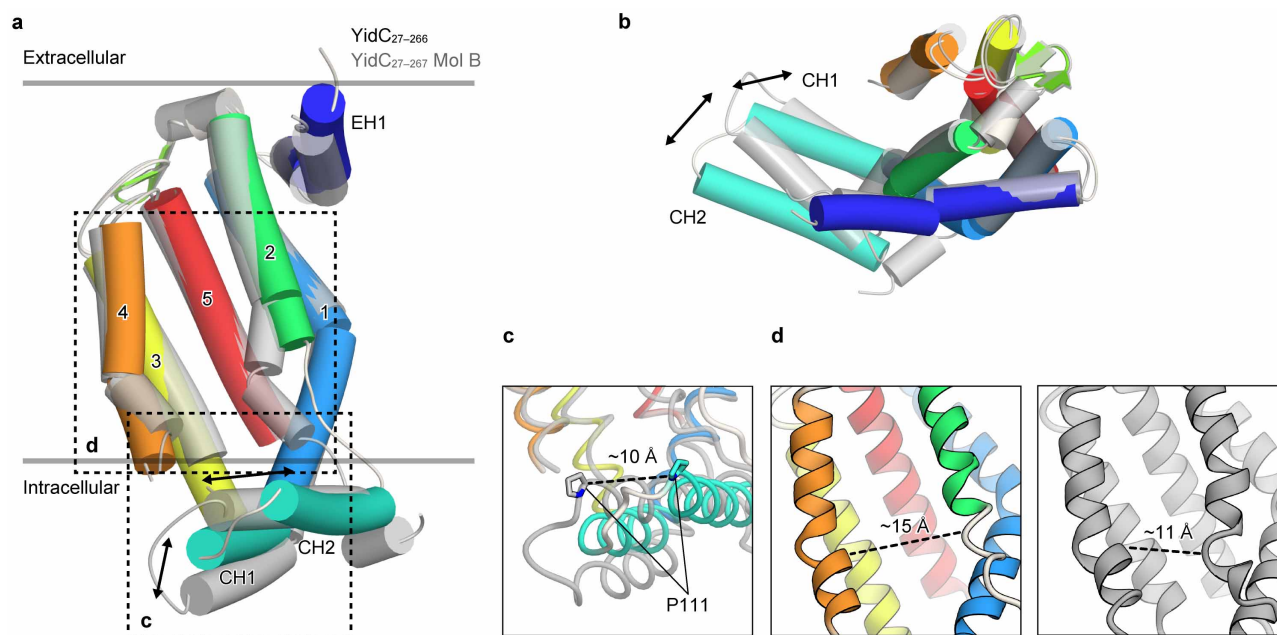
Extended Data Figure 2 | Electron density map of *B. halodurans* YidC. Stereo view of the $2mF_O - DF_C$ electron density map of the TM2 helix, contoured at 1.1σ , where m is the figure of merit and D is the SIGMA-A

weighting factor. F_O and F_C are the observed and the calculated structure factor amplitudes, respectively.



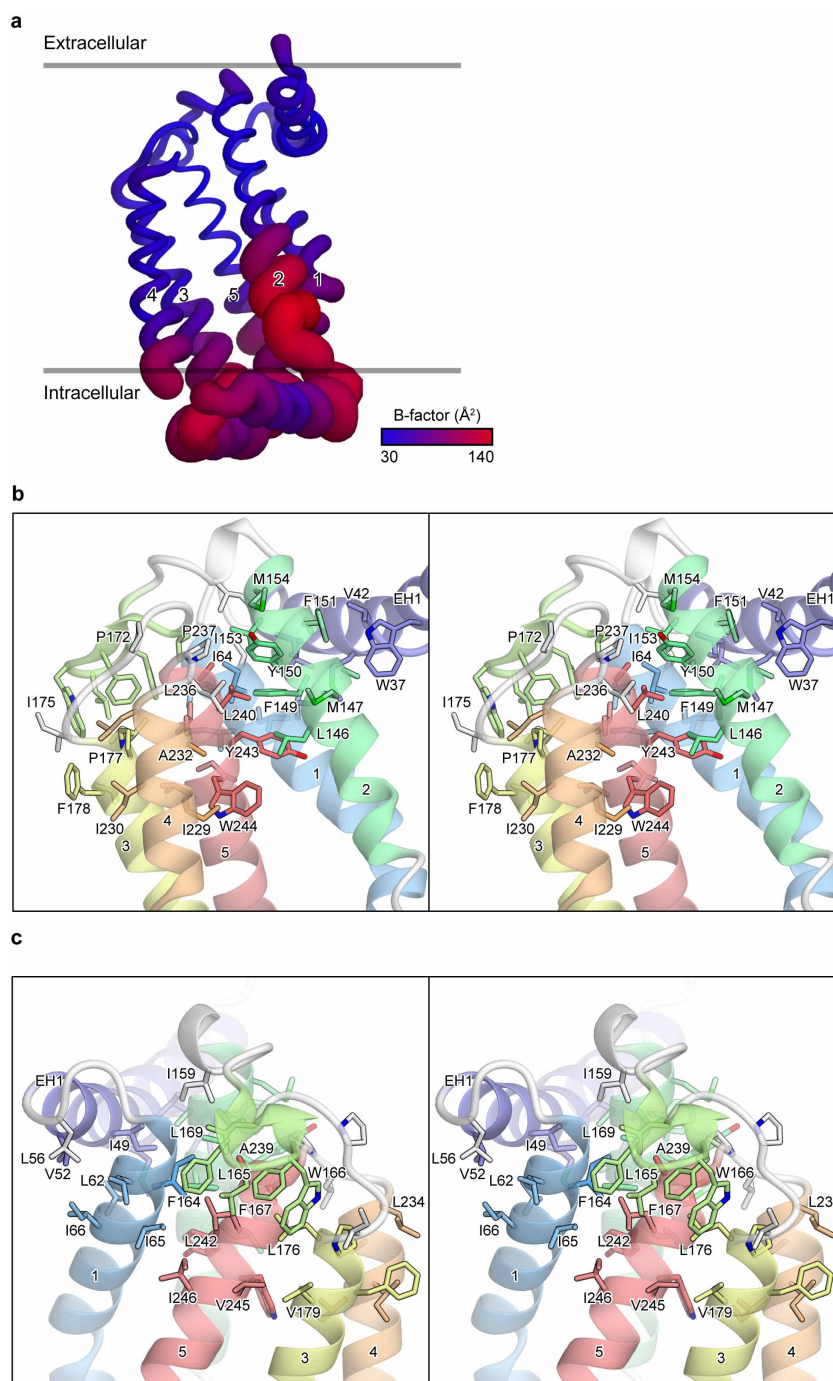
Extended Data Figure 3 | Monomeric *B. halodurans* YidC. **a**, The crystal packing of YidC_{27–266}, viewed from the plane of the membrane. The molecule in the asymmetric unit is coloured red. **b**, The crystal packing of YidC_{27–267}, viewed from the plane of the membrane. Two molecules (Mol A in light pink and Mol B in light blue) are in the asymmetric unit. **c**, The chromatograms show the ultraviolet (UV), refractive index (RI) and light scattering (LS) detector readings. The volume delays of UV between the other detectors were corrected. The traces were normalized to the peak maxima. The green and blue

lines in the LS chromatogram indicate the calculated molecular masses of the protein–detergent complex and the protein, respectively. **d**, The RI of the mixture was measured in response to five concentration steps. The refractive index increment (dn/dc) of the mixture of DDM and CHS was determined using linear regression of the RI versus the concentration. **e**, The molecular mass values determined by SEC-MALLS and calculated from the amino acid sequence.



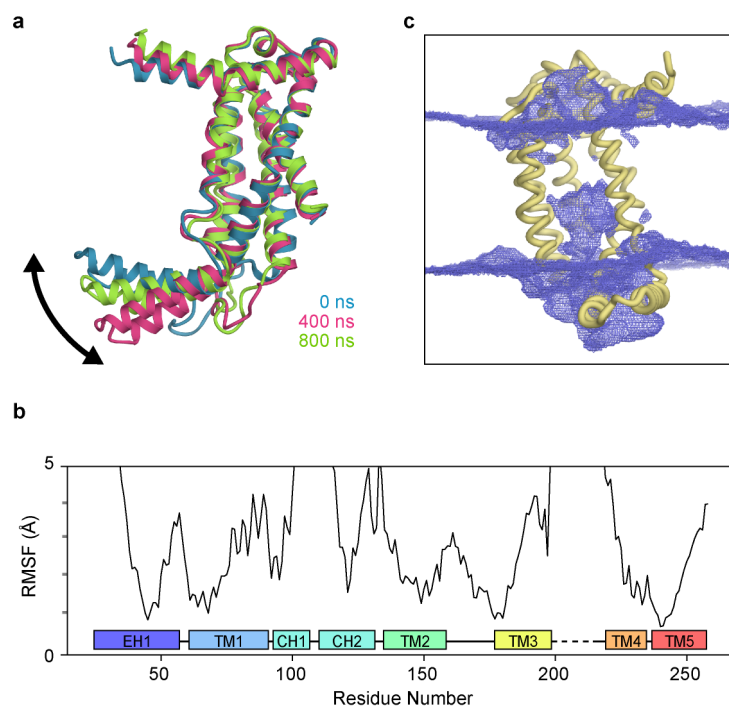
Extended Data Figure 4 | Structural flexibility of the hydrophilic groove and C1 region. **a, b**, Superimposition of the crystal structures of YidC₂₇₋₂₆₆ (coloured) and Mol B of YidC₂₇₋₂₆₇ (grey), viewed from the plane of the membrane (**a**) and from the extracellular space (**b**). The conformational changes observed in CH1 and CH2 are indicated by black arrows. **c**, Close-up view of the C1 region. The side chains of P111 are shown by stick models. In the

YidC₂₇₋₂₆₇ structure, the arrangement of the C1 region with respect to the core region is rotated by $\sim 35^\circ$ compared with that in the YidC₂₇₋₂₆₆ structure. As a result, the tip of the C1 region is displaced by ~ 10 Å in the YidC₂₇₋₂₆₇ structure. **d**, Close-up views of the hydrophilic groove (left, YidC₂₇₋₂₆₆; right, Mol B of YidC₂₇₋₂₆₇). The distances between the C α atoms of C136 and M221 are indicated by dashed lines.



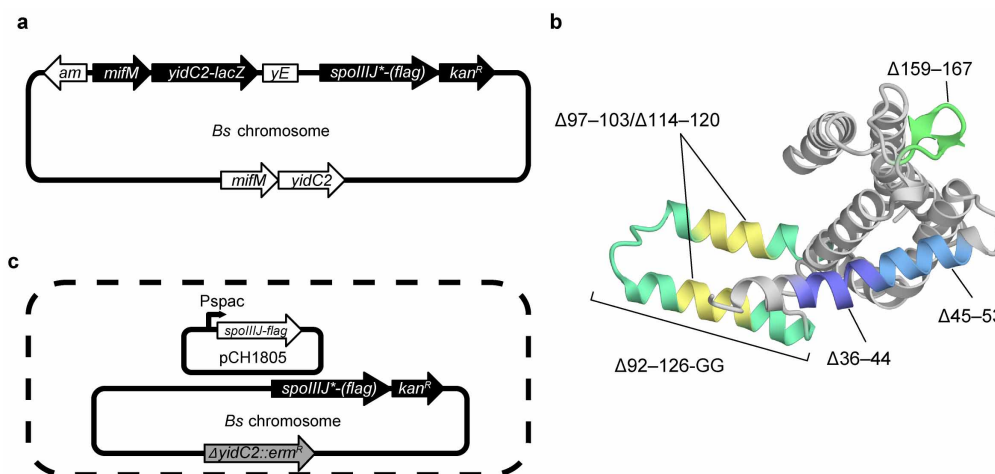
Extended Data Figure 5 | The hydrophobic core of *B. halodurans* YidC.
a, The crystallographic B-factors are coloured in a gradient varying from blue to

red, representing 30 to 140 \AA^2 . **b**, **c**, Stereo views of the hydrophobic core, showing the side chains of the hydrophobic residues.



Extended Data Figure 6 | Molecular dynamics simulation of *B. halodurans* YidC for 1,000 ns in a lipid bilayer. **a**, Snapshots of the structure over the time course of the simulation at 400 ns intervals: 0 ns (blue), 400 ns (magenta) and 800 ns (light green). **b**, Root mean square fluctuation (r.m.s.f.) of YidC during

the simulation. The secondary structure of YidC is indicated below the line. **c**, The water probability density map in the simulation, contoured at 0.001 molecules Å⁻³ ns⁻¹.



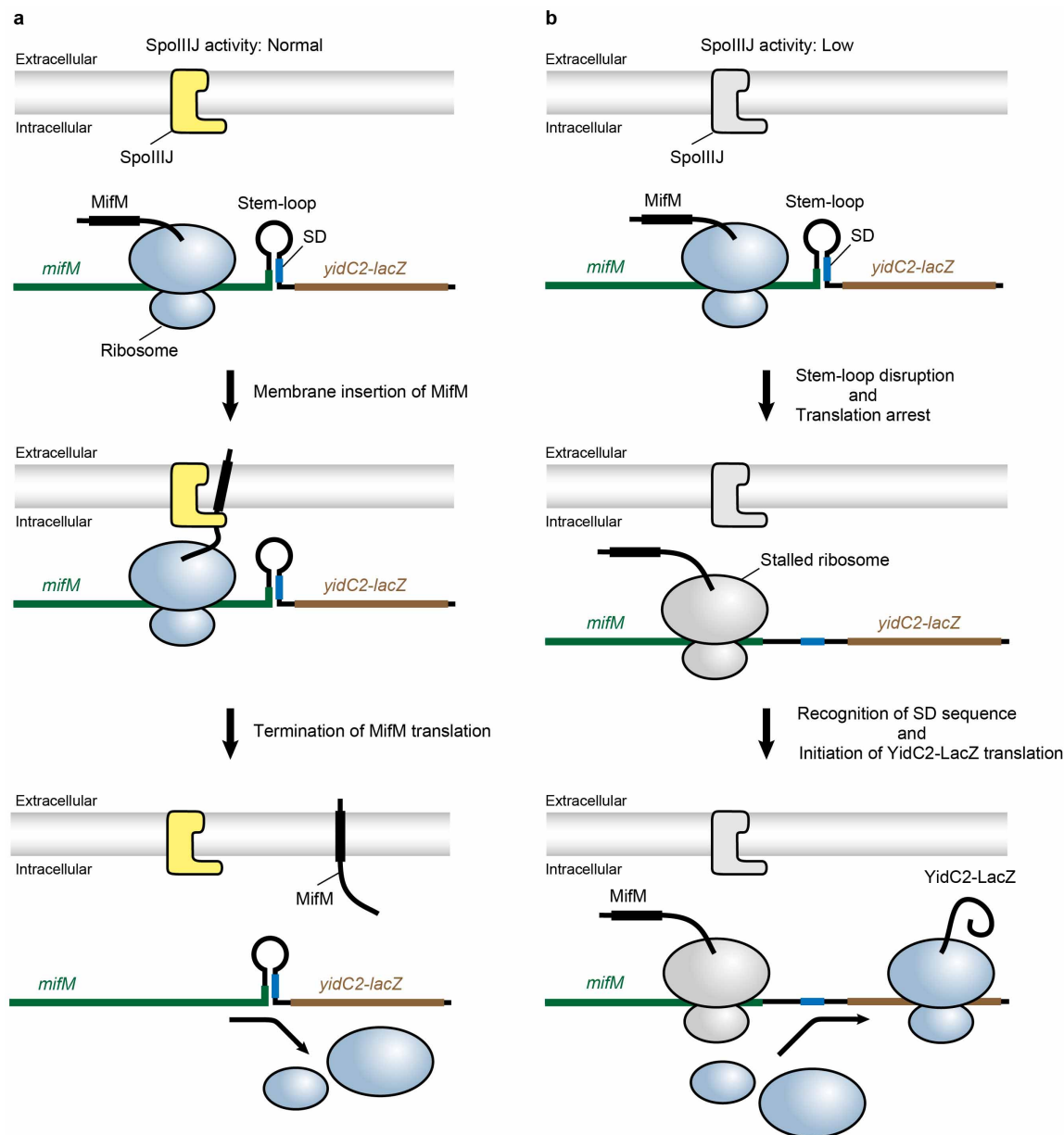
Extended Data Figure 7 | Gene structures of the strains and YidC mutants used for *in vivo* genetic analyses.

a, Schematic representations of the gene structures of the *yidC2-lacZ* reporter strains used for the MifM-based assay. *spoIIIJ*-(flag)* indicates either wild type or mutant *spoIIIJ*. *yidC2-lacZ* represents a translational gene fusion with the *lacZ* sequence in-frame after the sixth codon of *yidC2*. The native *mifM-yidC2* on the chromosome remained intact.

b, Deleted regions of SpoIIIJ, viewed from the extracellular side. Residue numbers in SpoIIIJ are indicated. Δ92–126-GG represents a mutant in which the entire C1 region has been replaced by a glycine–glycine linker. Δ97–103/Δ114–120 represents a mutant in which both the CH1 and CH2 helices have been shortened by seven residues.

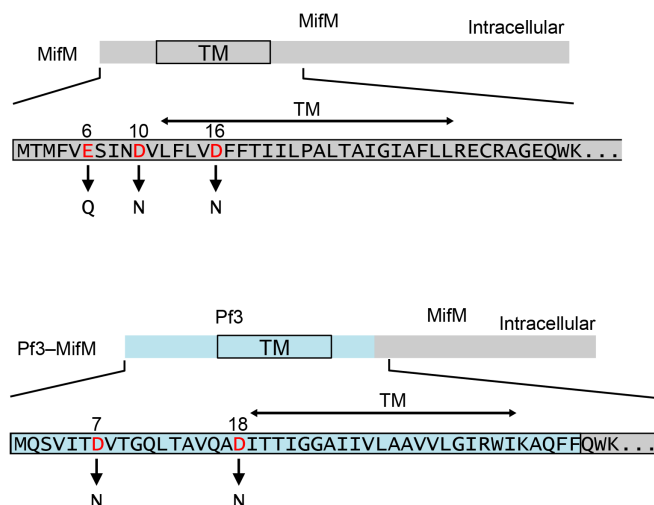
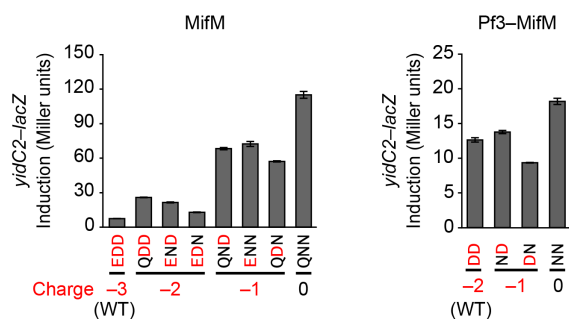
c, Schematic representations of

the gene structures used for growth complementation assays. SpoIIIJ becomes essential for the growth of *B. subtilis* when *yidC2* is disrupted. Cells with a disruption of chromosomal *yidC2* were transformed with the rescue plasmid pCH1805, which expresses wild-type *spoIIIJ-flag* under the control of the IPTG-inducible Pspac promoter. The native *spoIIIJ* on the chromosome was replaced by either wild-type or mutant *spoIIIJ* (*spoIIIJ*-(flag)*). In the absence of IPTG, *spoIIIJ-flag* is not expressed from the plasmid, making the chromosomal *spoIIIJ*-(flag)* the only source of cellular YidC. The complementation test measures the global role of SpoIIIJ in inserting a wide range of membrane proteins, including single-spanning and multi-spanning membrane proteins.



Extended Data Figure 8 | Schematic explanation of β -galactosidase activity assay and MifM insertion activity. **a, b**, MifM is a single-spanning membrane protein, and its membrane insertion is considered to be mediated by YidC (SpoIIIJ)²¹. To evaluate the MifM insertion activity of SpoIIIJ, we performed a genetic analysis using *B. subtilis*. In *B. subtilis*, SpoIIIJ is constitutively expressed, whereas YidC2 is expressed only when the SpoIIIJ activity is compromised, by the following mechanism. The expression of *yidC2* is regulated by the upstream *cis* regulator open reading frame of *mifM*, which is co-transcribed with *yidC2*. During the synthesis of MifM, the C-terminal region of nascent MifM interacts with the peptide exit tunnel of the ribosome and causes translational arrest^{40,50}. When the SpoIIIJ activity is normal, the

translational arrest is released by the SpoIIIJ-dependent membrane insertion of MifM. Therefore, the translational arrest is transient or does not occur (**a**). By contrast, when SpoIIIJ activity is compromised, MifM is not inserted into the membrane, and its translation is arrested, which causes ribosome stalling. The stalled ribosome disrupts the downstream stem-loop structure and exposes the Shine-Dalgarno (SD) translation initiation signal sequence of the *yidC2* messenger RNA (**b**). Thus, we can estimate the *in vivo* SpoIIIJ activity by measuring the expression of the introduced *yidC2-lacZ* fusion (Extended Data Fig. 7a); the reduction of MifM insertion efficiency by SpoIIIJ elevates the LacZ activity^{21,50}.

a**b**

Extended Data Figure 9 | Effects of N-terminal negatively charged residues of substrates on insertion. **a**, Schematic representations of the N-terminal negatively charged residues of MifM and the Pf3-MifM chimaeric protein.

b, Membrane insertion efficiencies of MifM mutants and Pf3-MifM mutants. The efficiencies were determined by the LacZ activities (mean \pm s.d., $n = 3$). The N-terminal negatively charged residues of MifM and Pf3-MifM and the numbers of the charged residues are shown at the bottom (EDD, wild-type MifM; DD, wild-type Pf3-MifM). Mutations of the acidic residues in the Pf3 coat protein had less pronounced effects than mutation of those in MifM, probably because the membrane insertion is facilitated by multiple interacting factors depending on the amino acid sequence.

Extended Data Table 1 | Data collection and refinement statistics

	YidC27–266	YidC27–267	YidC27–266 (Y150C) (Methyl HgCl)	
Data collection				
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.9, 60.6, 58.9	54.0, 70.1, 83.1	43.8, 59.7, 58.6	
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 100.3, 90	90, 92.0, 90	90, 100.3, 90	
			<i>Peak</i>	<i>Inflection</i>
Wavelength	1.000	1.378	1.00000	1.00945
Resolution (Å)	50–2.40	45–3.20	50–3.00	50–3.00
	(2.44–2.40)	(3.40–3.20)	(3.05–3.00)	(3.05–3.00)
<i>R</i> _{sym}	0.068 (0.318)	0.110 (0.472)	0.089 (0.350)	0.101 (0.456)
<i>I</i> / <i>σ</i> <i>I</i>	22.5 (2.85)	10.1 (2.96)	29.3 (4.72)	21.5 (2.75)
Completeness (%)	91.0 (85.6)	95.6 (99.4)	99.3 (98.3)	99.2 (98.3)
Redundancy	2.6 (1.9)	3.9 (3.8)	5.8 (4.4)	5.4 (3.5)
Refinement				
Resolution (Å)	2.40	3.20		
No. reflections	10,879	19,127		
<i>R</i> _{work} / <i>R</i> _{free}	24.2/25.9	26.0/29.1		
No. atoms				
Protein	1,601	3,586		
Ligand/ion	60	4		
Water	8			
B-factors				
Protein	65.24	96.2		
Ligand/ion	77.73	121.0		
Water	49.43			
R.m.s deviations				
Bond lengths (Å)	0.002	0.002		
Bond angles (°)	0.714	0.765		
Ramachandran plot				
Favoured (%)	99.5	96.1		
Allowed (%)	0.5	3.6		
Outliers (%)	0.0	0.2		

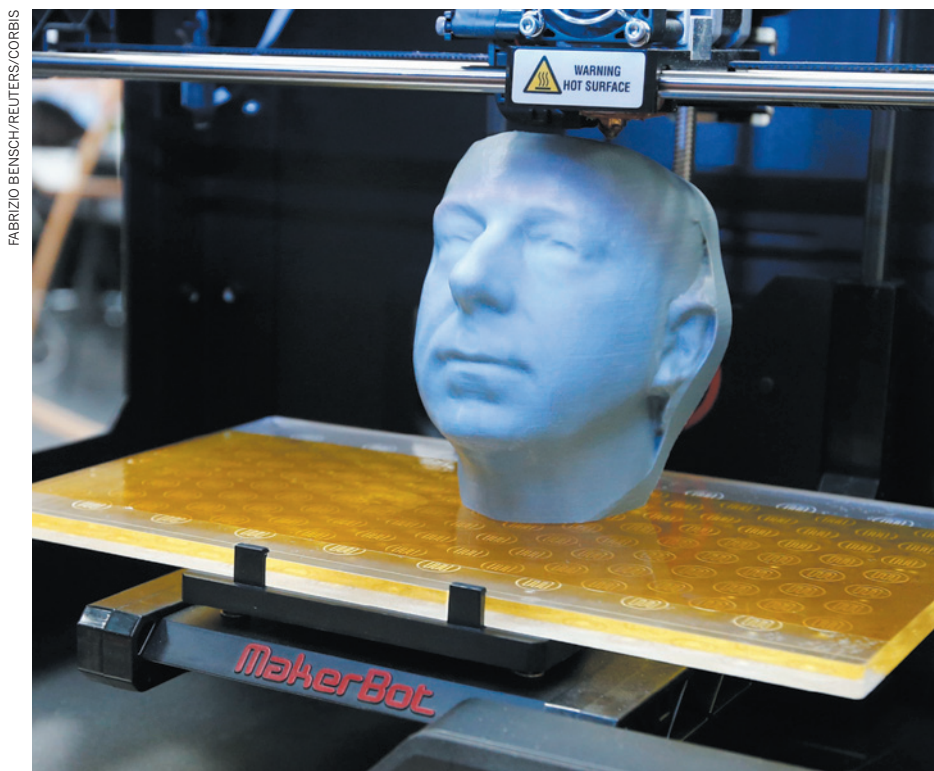
*The numbers in parentheses are for the highest resolution shell

CAREERS

RESEARCH EFFICIENCY Administrative tasks gobble scientists' time, says report **p.523**

AWARDS Prize winners invest their money in postdoc endowment **p.523**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



additive manufacturing, is a growing field that draws researchers from a variety of disciplines — computing, materials science, mechanical and electrical engineering, physics, chemistry, and even biology. They are in demand to develop the technology into something that might change how manufacturing is done and create all sorts of products that have not been possible before, from lighter-weight aircraft parts to cubes that unfold into furniture to 'biobots' made with living cells.

Wohlers Associates, a consulting firm in Fort Collins, Colorado, that focuses on the industry, says that the worldwide market for 3D printing products and services jumped to US\$3.1 billion in 2013 (a rise of 35% from 2012), and expects it to grow to \$10.8 billion by 2021. That will certainly open up more jobs in the field, says company president Terry Wohlers. "As the industry grows, so will the need for designers, scientists and researchers — especially chemists — to develop new materials," he says. "Also, we'll need mechanical and electrical engineers to develop new-generation 3D printers, manufacturing engineers to develop processes and supply chains and educators to introduce all of this to a new generation of designers, technicians and others."

Some of those jobs will be in the 30 or so companies that manufacture the printing systems in Europe, the United States, China and Japan. "The prospects of employment are very good," says Cathy Lewis, chief marketing officer at one of those companies, 3D Systems in Rock Hill, South Carolina, which lists numerous openings for engineers of different stripes. She says it's hard to project exact numbers of future hires, but points out that the company has grown from 300 employees four years ago to 1,500 today, many of them in technical positions.

Governments around the world are pouring money into research projects intended to advance the technology of 3D printing from a niche field to a major part of manufacturing, providing funds to encourage partnerships between universities and industry. In 2012, for instance, the US government launched the National Additive Manufacturing Innovation Institute, since dubbed America Makes, based in Youngstown, Ohio, with \$30 million in government funding and another \$40 million from industry. The institute has given out \$13.5 million in matching funds to research teams from both industry and academia. The US National Science Foundation, the Defense Advanced Research Projects Agency and NASA are all also funding 3D printing research. ►

Three-dimensional printing is opening up applications in many areas, including research.

TECHNOLOGY

Building opportunities

More jobs and research positions are being created as interest in 3D printing grows.

BY NEIL SAVAGE

Kiril Vidimčė began his professional life as a software graphics engineer in California — he spent eight years at Pixar Animation Studios in Emeryville and another four at chip-maker Intel in Santa Clara. But he wanted to move beyond the world of animated movies and video games into something that existed outside of a video screen. So in February 2012, he joined the Computational Fabrication Group at Massachusetts Institute of

Technology in Cambridge. He's now pursuing a PhD in computational fabrication and writing software that tells three-dimensional (3D) printers how to build complicated objects.

"I'm excited about working on physical things for a change. I've never been a maker myself, but I always wanted to be. I wanted to play with more physical stuff," Vidimčė says. "I think it's fun to get to the intersection where you're combining knowledge of computer science and electronics and materials."

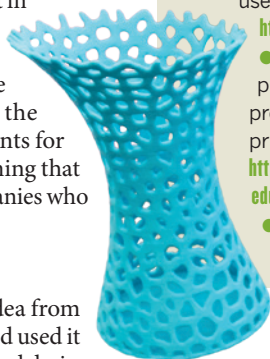
Indeed, 3D printing, also known as

► In 2013, the Singapore government announced that it would invest S\$500 million (US\$400 million) over five years in research into 3D printing and other innovative technologies such as robotics, and the Chinese government intends to invest \$240 million in 3D printing over three years. Last summer, the UK's Technology Strategy Board and the Research Councils announced £8.4 million (US\$14.22 million) in such funding. It is hard to say exactly how many jobs all that translates into, but David Bourell, a mechanical engineer and materials scientist who runs the Laboratory for Freeform Fabrication at the University of Texas, Austin, says he hears of universities hiring more faculty who focus on 3D printing. "There is an upswing in the past few years as public and governmental interest in the field has increased," he says. "That's the topic they want to hire in." In part, he says, the surge in interest is being driven by the expiration of the founding patents for 3D-printing technologies, meaning that it is now less expensive for companies who want to get involved in the field.

VALUE ADDED

For his part, Vidimče took an idea from computer-graphics rendering and used it to create OpenFab, computer-aided design software for building an object from more than one material. A computer model describing the 3D shape of an object, as well as the specific areas where different materials go, would be so large that it could halt the printing process. So instead of trying to model the whole object at once, Vidimče's programme breaks it up into small sections, and provides the printer with a recipe for what to do in each section. That is the same way that graphics software renders complex images in a video game, treating them as a series of small problems instead of one large one.

At 3D Systems, Mehdi Mojdeh, director of 3D printing and a technical fellow, says that much of the company's research is aimed at improving the speed and accuracy of the process. Although industry has been using the



- 3D Systems in Rock Hill, South Carolina, offers internships at various locations, including New York, Los Angeles and Boston, for people thinking about pursuing research in 3D printing to get some first-hand experience with the machines. <http://www.3dsystems.com>
- Many universities have 3D printers in one research lab or other. And hackerspaces, which provide the public with various types of equipment to build things, usually have one, and will often offer training in how to use it. <http://hackerspaces.org/wiki>
- The website 'Rapid Today' provides a list of degree programmes that include 3D printing. http://www.rapiddtoday.com/get_prototyping_education.html
- The online *Rapid Prototyping Journal* gives an entry into the

CASE STUDY

Learning about 3D printing

literature on current and past research in 3D printing.

<http://www.emeraldinsight.com>

● Students could consider attending conferences to hear reports of current research and to talk to people in the field.

● The Additive Manufacturing and 3D Printing International Conference in Nottingham, UK, on 7–11 July.

<http://www.am-conference.com/>

● The 25th Annual International Solid Freeform Fabrication Symposium in Austin, Texas, on 4–6 August.

<http://sffsymposium.engr.utexas.edu/>

● The Inside 3D Printing Conference and Expo is a business-to-business trade show with several shows a year around the globe.

<http://www.mediabistro.com/inside3dprinting/>

● The Rapid Prototyping Calendar lists a variety of upcoming conferences.

<http://www.additive3d.com/conf.htm> N.S.

technology for a while to design and prototype new products, "we are trying to push 3D printing more into a manufacturing and production environment," he says.

General Electric (GE) in Fairfield, Connecticut, is increasingly using the technology for building parts that go into its products, from refrigerators to jet engines. The company has more than 500 employees working directly on 3D printing, and Stephan Biller, GE's chief manufacturing scientist, says that this is likely to increase. "There's no doubt about it, additive manufacturing is going to change GE and it's going to change manufacturing as a whole," Biller says. "I don't think people will have trouble finding a job in this area."

One of the greatest needs, Mojdeh says, is for researchers who focus on the properties of materials. Companies would like to expand the range of material properties such as strength, hardness, conductivity and so on, that are available. The wider the range of properties, the more types of objects that can be built, and the more functions such objects can have. Another key need is for scientists who work on an ultimate goal of the industry: multi-material printing, or the building of devices made of different substances — say, a ceramic engine part with metal circuitry built in. Very few products, after all, consist of only a single material, but today's 3D machines can print with only two different polymers at most.

Those goals mean that the industry needs both researchers who can develop materials with desirable characteristics, and those that can make the various printing technologies

work with those new materials — as well as with each other. It is not yet possible to print metal and plastic in the same machine, for example, because the heat needed to fuse the metal would vaporize the plastic. To find solutions, researchers from different science and engineering disciplines will have to work together, because an advance in one area depends on — and affects — others. "If you continually develop the materials, then processes will have to change, and design systems will have to evolve, so it's very much a moving research target," says Christopher Tuck, deputy director of the UK Engineering and Physical Sciences Research Council's Centre for Innovative Manufacturing in Additive Manufacturing at the University of Nottingham, UK.

In February, the centre launched a new post-graduate degree programme, part of a growing movement to train researchers to aim at those targets (see 'Case study'). It also aims to prepare people from disparate backgrounds to apply their expertise to 3D problems — say, a biologist wanting to print artificial organs with living cells, or a polymer chemist creating a more durable plastic part. The centre is looking for 66 students for a PhD programme that begins in October. "We're going to be recruiting scientists and engineers from much more outside mechanical engineering. It's going to be a lot of physics, chemistry, maybe a biologist," says Tuck.

All the students will spend their first year at Nottingham, getting a grounding in how 3D printing works, but they'll then specialize at one of four institutions, depending on their area of interest. Nottingham will focus



Jewellery makers benefit from 3D printing.

STEFANO TINTI/SHUTTERSTOCK

3D SYSTEMS

on design, multi-material printing and polymers. Loughborough University will deal with hybrid printing technologies. The University of Liverpool will focus on research into metals, and Newcastle University will specialize in additive manufacturing for biomedical uses.

GOING GLOBAL

Last September, Nanyang Technological University in Singapore opened the NTU Additive Manufacturing Centre. The centre offers master's degrees in precision engineering, mechanical engineering and manufacturing systems and engineering, all with a specialization in additive manufacturing, as well as a PhD in mechanical engineering with a topic in additive manufacturing. It is also providing internships to students from around the world.

The creation of the centre has led to postdoc and faculty positions in several disciplines, including

materials science, mechanical engineering, chemical engineering and biomedical engineering, says Chua Chee Kai, who heads it. He says it is too early to predict how many jobs related to 3D printing there might be in the business world in general, because the technology is in the early stages of adoption. "We do, however,

firmly believe that these are high-value jobs that will change the landscape of that particular business or sector," he says. And as the field grows, he predicts, "we will need not just engineers, but also computer experts, mathematicians, software engineers, material scientists, quality experts, system-solution providers and many more to sustain the economic ecosystem of 3D printing."

Bourell says that even institutions without formal programmes usually have some research going on. "You can pick almost any university and they have some kind of activity these days," he says. "They have at least one person dabbling." He expects the field to evolve and expand rapidly enough over the next 5 to 10 years for there to be considerable demand for people with master's- and PhD-level education. "We're right on the edge of the wave," Bourell says. "Now it's time to jump in." ■

Neil Savage is a freelance writer in Lowell, Massachusetts.



"If you continually develop the materials, then processes will have to change."

Christopher Tuck

RESEARCH EFFICIENCY

Time wasted

US scientists are spending 42% of their time on bureaucratic chores, according to a US National Science Board (NSB) report. The NSB named financial and personnel reporting and over-detailed preliminary grant proposals as common complaints. Arthur Bienenstock, chair of the NSB task force that authored the report, says that the need to submit a detailed budget in the first round of a grant proposal, for example, is 'nonsensical'. The NSB found that an over-regulation culture is overwhelming US scientists, and recommends limiting proposal requirements to those essential for evaluating merit; keeping reporting focused on outcomes; and automating payroll certification.

AWARDS

Postdoc prizes

Three science-award recipients based in New York City have pledged US\$3 million of their prize money to set up an endowment to fund awards for postdoctoral researchers. Charles Sawyers of Memorial Sloan Kettering Cancer Center, Cori Bargmann of Rockefeller University and Lewis Cantley of Weill Cornell Medical College last year each won \$3-million 'Breakthrough' prizes. The new Tri-Institutional Breakout Awards for Junior Investigators will offer three to six \$25,000 awards per year. The endowment founders say they are donating part of their prizes because they believe that postdocs will benefit more from the support. "For me, a prize at that stage of my career would have been really helpful," says Bargmann, who adds that postdocs do much of the most creative work in science. Inaugural winners will be announced by the end of the year.

MULTIPLE DISCIPLINES

Converging futures

Exciting future research opportunities will arise where disciplines converge — but government funding and academic structures are ill prepared to take advantage of them, says a report from the US National Research Council. Report committee chairman Joseph DeSimone says that this convergence will require scientists to understand the language of multiple disciplines and sectors, and calls 3D printing (see page 521) a 'wonderful example', as it depends on so many different disciplines.

TRIALS

An experimental direction.

BY JOHN FRIZELL

“You can’t do experiments on my boyfriend,” said Ellie. “I absolutely forbid it. It must be illegal, or unethical, or something.”

“He volunteered,” said Jamie. “He gets paid and all legal requirements are being met.”

“Does Mum know how much you’re spending?”

Although her little brother was a millionaire thanks to his inventions and patents, their mother controlled the money until he was 18.

“She won’t mind.”

“I love Claude. You better not hurt him.”

“You can always get another one. You change them all the time.”

“So what is this about?” she said, as an alternative to slapping him.

“Invisible earphones.”

“Well that’s good.”

She liked the idea. No set of earphones ever made looked remotely cool. She loved listening to music but didn’t want to go around with earphones on all day looking like a loser.

“How will you do it?”

Jamie’s gifts did not include the ability to explain his ideas to normal people. There was something about magnetosomes, tiny things with iron crystals in them.

“They are the compass of homing pigeons,” said Jamie, and then immediately contradicted himself by saying that what birds used was smaller and more subtle. “It may not work, but the point is that if these things can align themselves with a magnetic field then we can use a magnetic field to move them and induce signals directly into nerves.”

“The point is that you want to inject tiny particles into Claude’s brain. Well you can’t.”

But he could. Ellie’s appeal to their mother was in vain.

“There’s nothing to worry about, dear,” she said. “Claude is just one of nine students from your university who signed up for the experiment. It’s being run by a licensed human testing organization and there’s a full review and risk assessment. Jamie showed me.”

Ellie nodded and smiled. She always smiled at her mother when she was trying to change her mind.

NATURE.COM

Follow Futures:

[@NatureFutures](#)

[go.nature.com/mtoodm](#)

Ellie had googled the firm and the individual doctors but found nothing she could use

to stop the trials from going ahead.

“But it’s a lot of money. Even Jamie says he doesn’t know if it will work.”

“Oh don’t you worry about that. The accountants will offset it against tax. It won’t cost anything to try and it might be a big success.”

There was no change in Claude. He had



always been given to acts of spontaneous generosity, and now that he had more money Ellie got more lovely surprises. But her brother was glum. The hidden collar magnets produced no sounds. Bigger magnets did nothing. The volunteers had even put their heads into some sort of super magnet. None of them heard induced sounds. The trials had failed. She texted Jamie an invitation for coffee so she could cheer him up.

She had Claude on her arm when they met. Jamie gave him the polite, unengaged smile that he gave to most people.

“Shall we go to Coffee Delight?” she said.

“Let’s go to Midnight Mocha.”

It was typical Jamie. CD was just down the block and it was bright, bustling and super

popular — everyone was there. Midnight Mocha was gloomy and usually empty. But it was what Jamie liked and she was doing this for him. She started down the well-worn route but Claude’s strong warm arm steered her away from it.

“Let’s go this way,” he said.

They took a funny route through the crisp autumn air, cutting through a park, walking on the grass instead of following the paths and then took a sharp turn at the end, not the shortest way but nice. Jamie paced beside them, messing with his mobile phone.

Ellie tried to cheer Jamie up over cappuccino, but he was even more introverted than usual. After one of Claude’s jokes Jamie looked up.

“When we walked here you were always heading 17° west of south,” said Jamie.

“Why did you do that?”

“Just walking,” said Claude. “At least we got here.”

He took the opportunity to launch into a hilarious story that left Ellie doubled up with laughter and even brought a smile to Jamie’s face. When they left Jamie paid for everything: coffees, cakes and refills.

“I can charge it to the project,” he said.

It was a first. She must have succeeded in cheering him up; her tight-fisted brother never paid for anyone else.

Two mornings later, tears were streaming down Ellie’s face as she tried to eat her low-fat yogurt breakfast. She wiped them away as Jamie came into the kitchen.

“Is Claude leaving you?”

She sat up, astonished. Jamie could hardly tell if other people liked or disliked him. He had never had any sort of relationship in his life and was hopeless at picking up the most basic social cues.

“How did you know?”

“The way he walked to Midnight Mocha. If you apply the magnetic declination to that heading, he was walking straight away from magnetic north.

“But Jamie! He’s quitting university. He’s going to Florida. I’ll never see him again!”

“Do you love him enough to wait for him?”

“Oh, Jamie. You know I do.”

“Then it will be OK. He’s just migrating. Don’t worry, he’ll be back in the spring. ■

John Frizell was trained in biochemistry and works in ocean conservation for Greenpeace. In his spare time he walks, builds robots and sings.

JACEY